Research Report 2004

Augmented Reality Mentor
Technical and Evaluation Report

**Rakesh Kumar, Supun Samarasekera, Girish Acharya,
Louise Yarnall, Zhi-Wei Zhu, Michael Wolverton,
Vlad Branzoi, Glenn Murray, Nicholas Vitovitch,
Ryan Villamil, & Jim Carpenter**
SRI International

July 2017

**United States Army Research Institute
for the Behavioral and Social Sciences**

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

SRI International

Technical review by:

William R. Bickley, PhD
Louis C. Miller, PhD

NOTICES

DISTRIBUTION: Approved for public release; distribution is unlimited.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* July 2017	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* September 2014-February 2016
4. TITLE AND SUBTITLE Augmented Reality Mentor Technical and Evaluation Report		**5a. CONTRACT NUMBER** W15QKN-13-C-0083 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER** 644775
6. AUTHOR(S) R. Kumar, S. Samarasekera, G. Acharya, L. Yarnall, Z. Zhu, M. Wolverton, V. Branzoi, G. Murray, N. Vitovitch, R. Villamil, & J. Carpenter		**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) SRI International 333 Ravenswood Avenue Menlo Park, CA 94025		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610		**10. SPONSOR/MONITOR'S ACRONYM(S)** ARI **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** Research Report 2004

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

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter POC: William R. Bickley, PhD

14. ABSTRACT
A prototype visual augmented reality (AR) system, designated "AR Mentor," for training maintenance on the U.S. Army Bradley fighting vehicle was developed and tested. The system consists of a compact computer, head worn cameras, microphone, ear-buds and eyewear. A virtual personal assistant provides real-time dialog and reasoning supporting human-like interaction using spoken natural language. Feedback and interaction occurs both verbally and by engaging the AR system to display icons and instructions visually on a monocular optical see-thru display. The inserted visual objects appear as part of the live scene and remain precisely aligned to the equipment. The prototype's potential for training was evaluated in the Ft Benning Bradley Training Division's introductory training course for Bradley maintainers. Even though the prototype's training capabilities were not optimized, student hands-on learning on two types of maintenance tasks while using the system was still equivalent to learning achieved under normal tutelage of an Army instructor.

15. SUBJECT TERMS
Augmented reality, maintenance training, Bradley maintenance, virtual personal assistant

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**Dr. Scott E. Graham**
Unclassified	Unclassified	Unclassified	Unlimited A		**19b. TELEPHONE NUMBER** 706-545-2362

Research Report 2004

Augmented Reality Mentor Technical and Evaluation Report

Rakesh Kumar, Supun Samarasekera,
Girish Acharya, Louise Yarnall, Zhi-Wei Zhu,
Michael Wolverton, Vlad Branzoi,
Glenn Murray, Nicholas Vitovitch,
Ryan Villamil, & Jim Carpenter
SRI International

Ft. Benning Research Unit
Scott E. Graham, Chief

United States Army Research Institute
for the Behavioral and Social Sciences

November 2016

Army Project Number
633007A792

Personnel Performance
and Training

AUGMENTED REALITY MENTOR TECHNICAL AND EVALUATION REPORT

EXECUTIVE SUMMARY

Research Requirement:

The Army Learning Concept (Department of the Army, 2011) re-orients Army training toward the individual Soldier and calls for increased reliance on technology to deliver that training. Augmented reality (AR) is an emerging technology with the potential of being integrated into an individualized training environment, but, in and of itself, AR is only a training medium with potential. One way to bridge the gap between AR potential training and AR mediated training would be to integrate with the AR an automated sequencing of instruction that uses AR at the appropriate points during the sequence. One way to implement the automated sequencing would be via a virtual personal assistant (VPA) capability that "personally" guides the Soldier trainee through the instruction.

The research reported here addresses the prototype development, integration, and assessment of a combined AR and VPA functionality (dubbed AR Mentor) implemented for U.S. Army Bradley infantry fighting vehicle maintenance training. The research addresses two issues: does the combined technology appear technically feasible, and does it appear to be of use in training individual Soldiers

Procedures:

In its final configuration, AR Mentor consisted of three physical components: a head mounted display (HMD) connected to a wearable processor which is wirelessly connected with a separate server. The HMD consisted of a monocular optical see-thru display, microphone, navigational camera, and inertial measurement unit. A battery pack and high-priority (e.g. navigation, rendering) processor were attached to a vest worn by the trainee, and the separate server handled lower priority (e.g. speech recognition, dialog control) processing.

With cooperation and assistance from the U.S. Army Maneuver Center of Excellence Bradley Training Division (BTD) cadre, AR Mentor's use was evaluated in two general areas: training for a straightforward maintenance task and training for troubleshooting procedures. Integrating VPA capability with visual AR, the system had the capability to "walk" a trainee thru any maintenance task. The VPA was implemented to mirror the exact same sequence a Bradley instructor would use for training the tasks.

Results:

Students in the Bradley maintainer's course were able to successfully use and operate AR Mentor. When students trained via AR Mentor were compared with students receiving normal BTD training, there was no appreciable difference between the two in terms of performance or learning.

Utilization and Dissemination of Findings:

AR Mentor was demonstrated to leaders and trainers at the U.S. Army Maneuver Center of Excellence and the U.S. Army Combat Arms Support Command.

AUGMENTED REALITY MENTOR TECHNICAL AND EVALUATION REPORT

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

AR Mentor Technical and Evaluation Report

Background

The prototype AR Mentor system was developed under a multi-year Army research effort aimed at characterizing augmented reality (AR) as a technology with applications to Army training. AR is a vision and visualization capability that enables the overlay of real-time visual information on a user's view of the physical world, to guide him or her in performing tasks. The capability relies on automated recognition of objects in the scene and precision localization of objects relative to the user's view. Inserted objects, icons and text appear as part of the live scene and appear to remain anchored to the scene as the user moves his or her head.

AR Mentor differs from typical AR applications in that it incorporates a virtual personal assistant (VPA) functionality. In this case, the VPA functionality comprises a real-time dialog and reasoning system supporting human-like interaction using spoken natural language. The system is based on automated speech recognition, natural language understanding, and reasoning. It is designed to recognize the user's goals and provide feedback to the user. The feedback and interaction occur both verbally and by engaging the augmented reality system to display icons and text visually on the user's viewing device. This functionality is similar to (and a successor to) widely used "intelligent interfaces" currently implemented on various smart devices.

AR Mentor was designed to train Soldiers to do maintenance and repair tasks for a variety of vehicles, weapons and complex machinery. The Army platform selected for this prototype was the M2A3 Bradley Infantry Fighting Vehicle. For the first phase, AR Mentor was used in training Soldiers to perform the Bradley "Tow Lift Limit Switch Adjustment" task for Bradley Fighting Vehicles. In the second phase, AR Mentor was extended to be used in training electronic diagnostic troubleshooting tasks for the Bradley. However, the system is designed to be general in its application and may be applied for maintenance or operation training of other equipment.

AR Mentor is a Soldier-worn augmented-reality mentoring system (Kumar 2014), and is configured to assist in maintenance, repair and diagnostic tasks of vehicles, weapons and complex machinery (Figure 1). It consists of user worn display eye-wear, microphone and head-phones configured to (a) talk to the Soldier to give directions, (b) display textual information of tasks and (c) overlay symbolic icons and directions that precisely align to the vehicles parts being observed. To do so it is important for the system to understand the task context. Context is obtained through (a) having a microphone fed speech understanding system that can listen to and interpret the Soldier's speech and (b) a video based sensor package that can accurately locate the Soldier with respect to the vehicle and interpret his actions. The system is hands-free and heads-up with natural spoken language interactions thus allowing trainees' uninterrupted attention to task while learning.

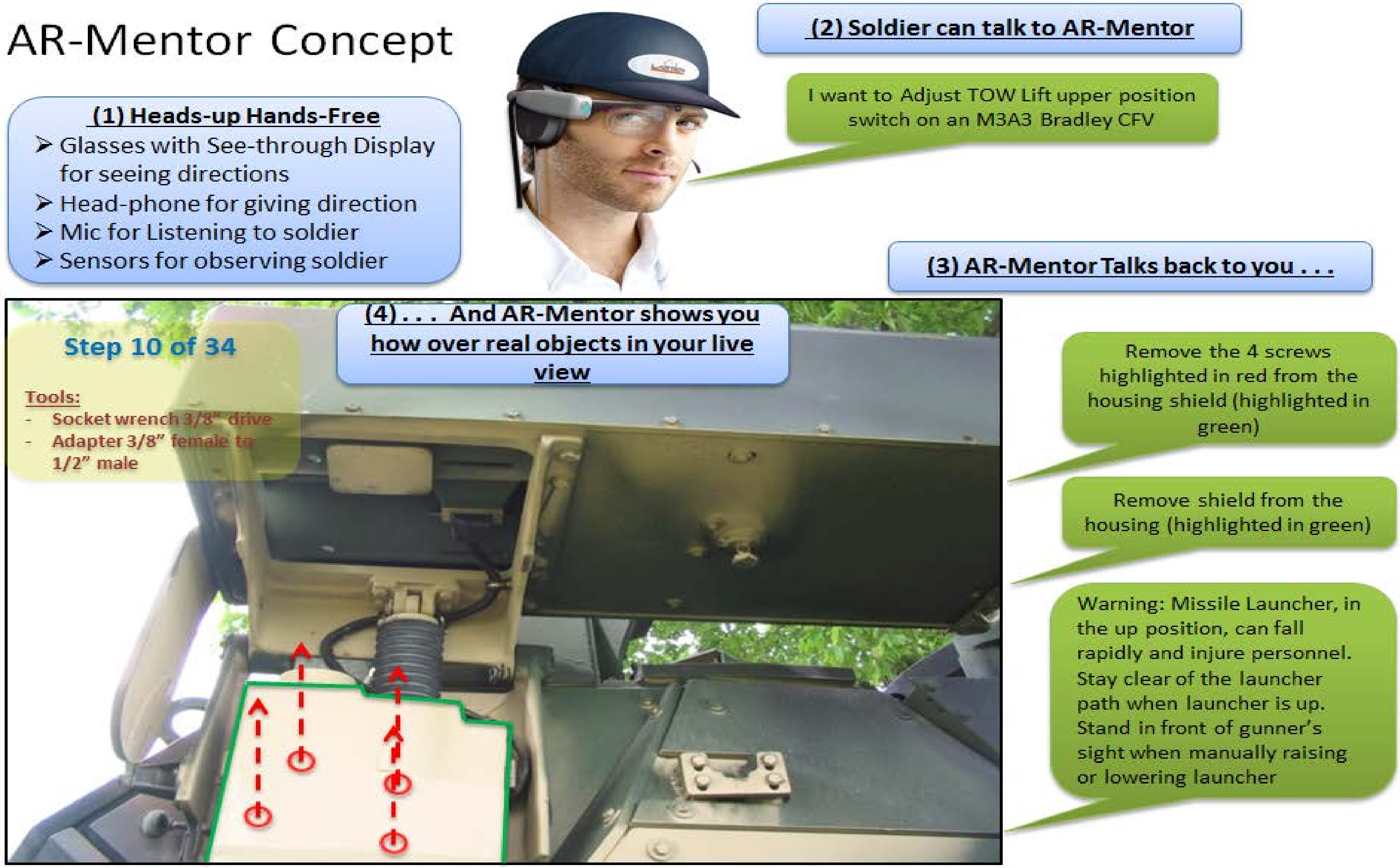

Figure 1: AR Mentor overall concept of operation.

This AR Mentor Technical and Evaluation Report covers the high level functional description of the AR Mentor system. It describes the system hardware and software components, their purposes, and their relationships. Finally, it presents an assessment and analysis of the system when used to train Soldiers for two tasks taught in the Bradley basic maintainer's course conducted by the Bradley Training Division at Fort Benning, GA.

Overall Software Architecture

Figure 2 shows in dark blue the key subsystems of the AR Mentor System. The sensor processing subsystem (SPS) interfaces with the trainee worn sensors. This includes the microphone to process trainee speech and Video/ Inertial Measurement Unit (IMU) based sensors to track the trainee's position, orientation and actions. The SPS block processes all the high-bandwidth, low-latency data to produce higher level information that is consumed by the down-stream sub-systems. The audio feed is converted to textual phrases. The video feed, along with the IMU data, is interpreted to find the trainees position with respect to the equipment and his gaze direction. The system also supports add-on modules for higher level constructs such as action recognition and object recognition.

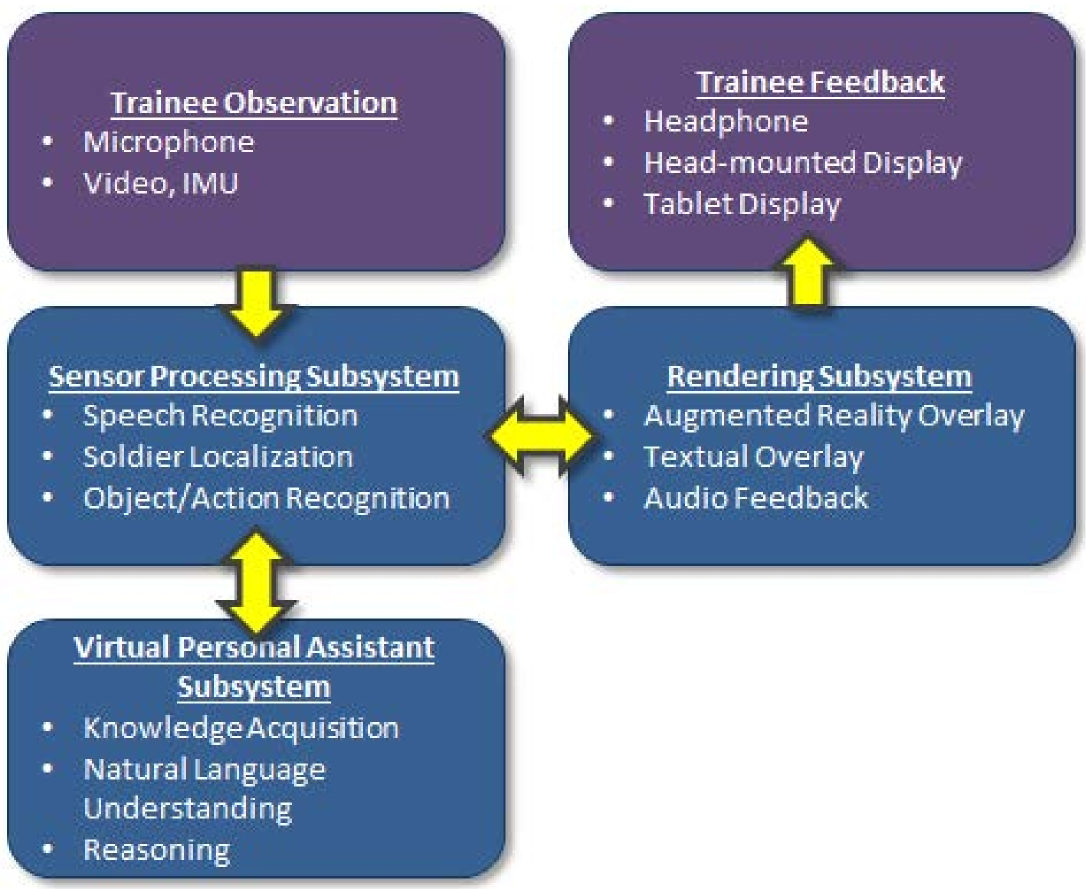

Figure 2. AR Mentor system block diagram

SPS coordinates interactions with the other two sub-systems: the rendering subsystem and the VPA subsystem. The VPA subsystem processes the higher level constructs from the SPS to construct a trainee intent. The intent is further analyzed using a knowledge base that represents the task workflow to generate interactive context that is generated by AR Mentor. VPA can also provide feedback to the SPS block on locales and actions of interest based on context. The VPA subsystem is setup as a stand-alone server that can be run remotely through low-bandwidth connections. The SPS takes directives from the VPA subsystem to instantiate specific detections of Interest.

The low-latency trainee location information generated by the SPS and VPA subsystems' interactions is forwarded to the rendering subsystem. The rendering subsystem generates AR overlay animations that exactly match the user's perspective view as overlays. VPA subsystem generated textual phrases are converted to speech for auditory feedback to the trainee.

Figure 3 shows the high-level distribution of software architecture across the hardware architecture.

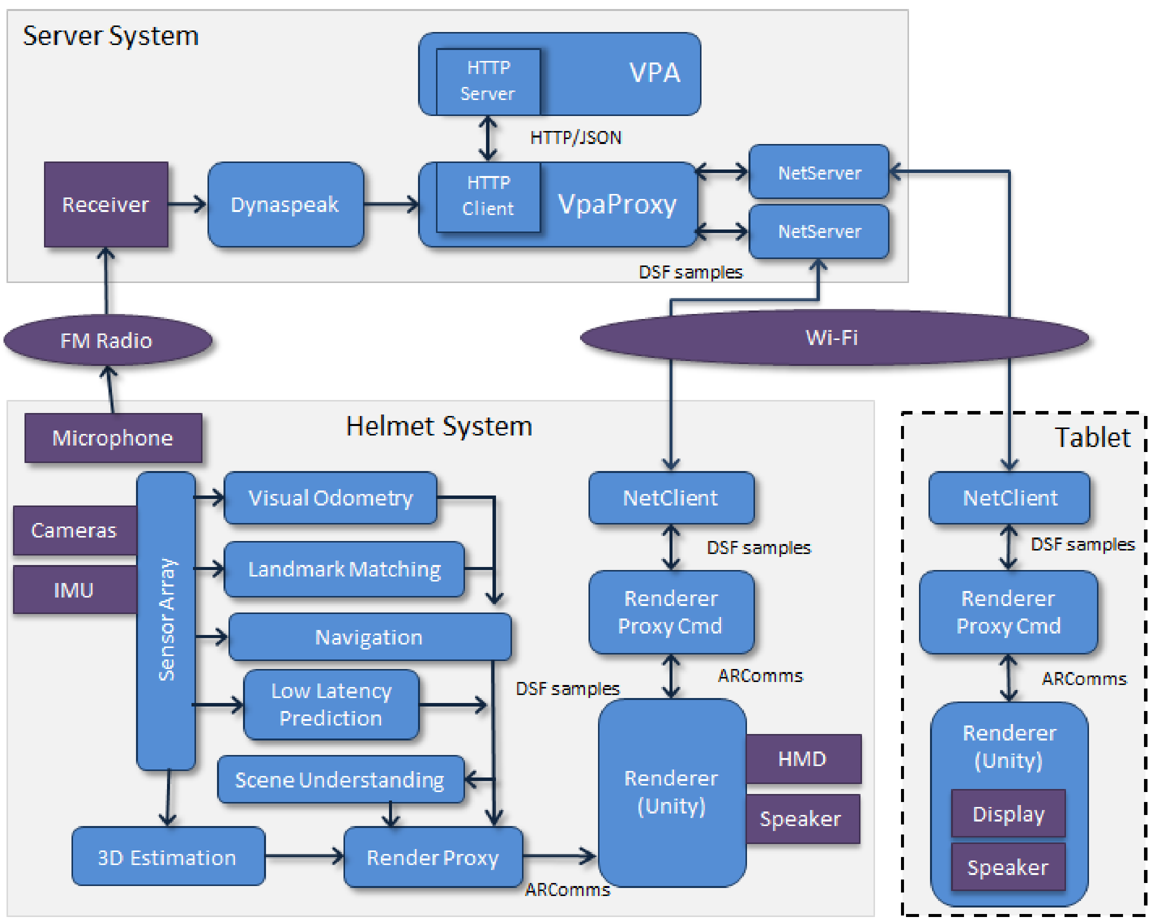

Figure 3. System components.

The SPS, VPA, and rendering subsystems are described in more detail below.

Sensor Processing Subsystem (SPS)

The SPS plays the central role of interfacing with all input sensors and the other subsystems. SPS also plays a key role in high-bandwidth data processing and low-latency feedback that is required by the AR Mentor system. A SRI developed data streaming framework (DSF) was utilized to support SPS. The DSF is a plug-and-play architecture that allows development of independent algorithm modules with specific streaming interfaces that can be connected at run-time without requiring software compilation. The DSF allows algorithm modules to be independently developed and interfaced using filters at run time for real-time data flow. These modules are discussed below.

SensorArray. The AR sensor hardware interfaces with the software through a unified I/O subsystem called SensorArray. This subsystem initializes, configures, and communicates with the sensors, pre-processes the sensor data, and synchronizes and packages sensor data for consumption by the other modules. Use of SensorArray enables changing, modifying and removing sensors without any other component being affected or needing to know anything about the underlying hardware.

This allows abstraction of different sensor API's enabling rapid reconfiguration for the sensors being used. Thus, upgrade to different hardware is possible while leaving the rest of the system unaffected.

DynaSpeak. DynaSpeak is a high accuracy, speaker independent, speech recognition engine that automatically adjusts to different speakers and accents. It also supports real-time dynamic noise compensation to handle background noise. For better speech-recognition performance, the ASR component was updated with language and acoustic models for the specific domain of repair of Army vehicles, weapons and machinery. Audio data was transcribed and annotated to build new, domain-specific speech models. A set of typical dialogs between the user and the system were captured to generate paraphrases (variations) for the dialogs and to create audio samples and built language and acoustic models.

Figure 4 illustrates the flowchart of speech processing. First, the speech signal is converted into a sequence of feature vectors. Phones (sounds) are modeled as a sequence of three-state Hidden Markov Models (HMMs). Each state represents a segment of a phone – beginning, middle, and end. Each state has associated a Gaussian mixture model to represent the acoustic features associated to that phone segment. Words are represented as probabilistic networks of phones. The language model provides probabilities to the different word sequences. The hierarchical structure can be flattened into a single large HMM by replacing the lower-order units into higher-order units.

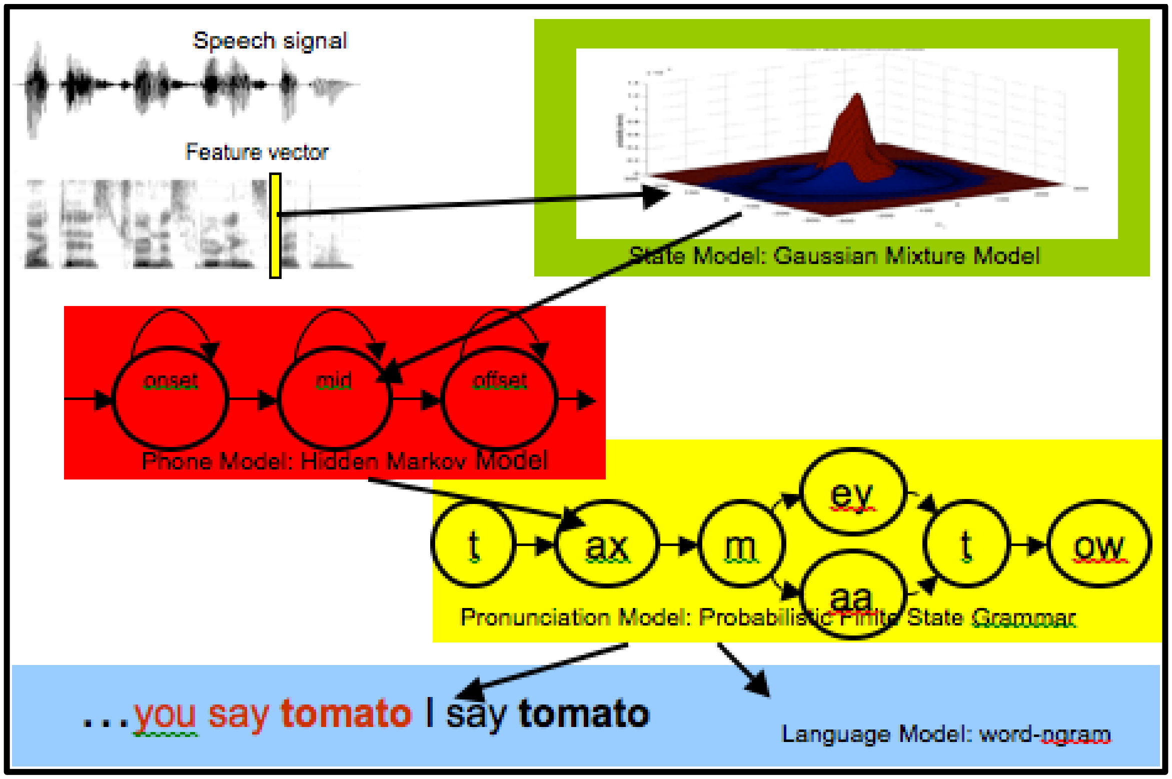

Figure 4. Speech processing flowchart.

Pose estimation. The goal of pose-estimation is to estimate the 6-DOF pose (3D location and 3D orientation) of trainee's eye-wear with respect to the vehicle or machinery under repair. The estimated pose is used to position the overlay of icons, symbology and annotations on the user's eyewear display. The inserted icons, symbols and annotations that are associated with points in the visual field (e.g. bolts, equipment covers) must not jitter or drift as the user moves his head, regardless of the rate at which the head moves. The overlay must also be accurate, with the correct items seen through the eyewear display annotated. In order to do this accurate and jitter/ drift-free overlay of icons, the AR Mentor system must estimate the pose of the eye-wear very accurately and with very low latency. This section describes the design of the different modules of the pose estimation filter (Figure 5) used to achieve the multiple goals of accuracy, low latency, no jitter, and drift free insertion of icons.

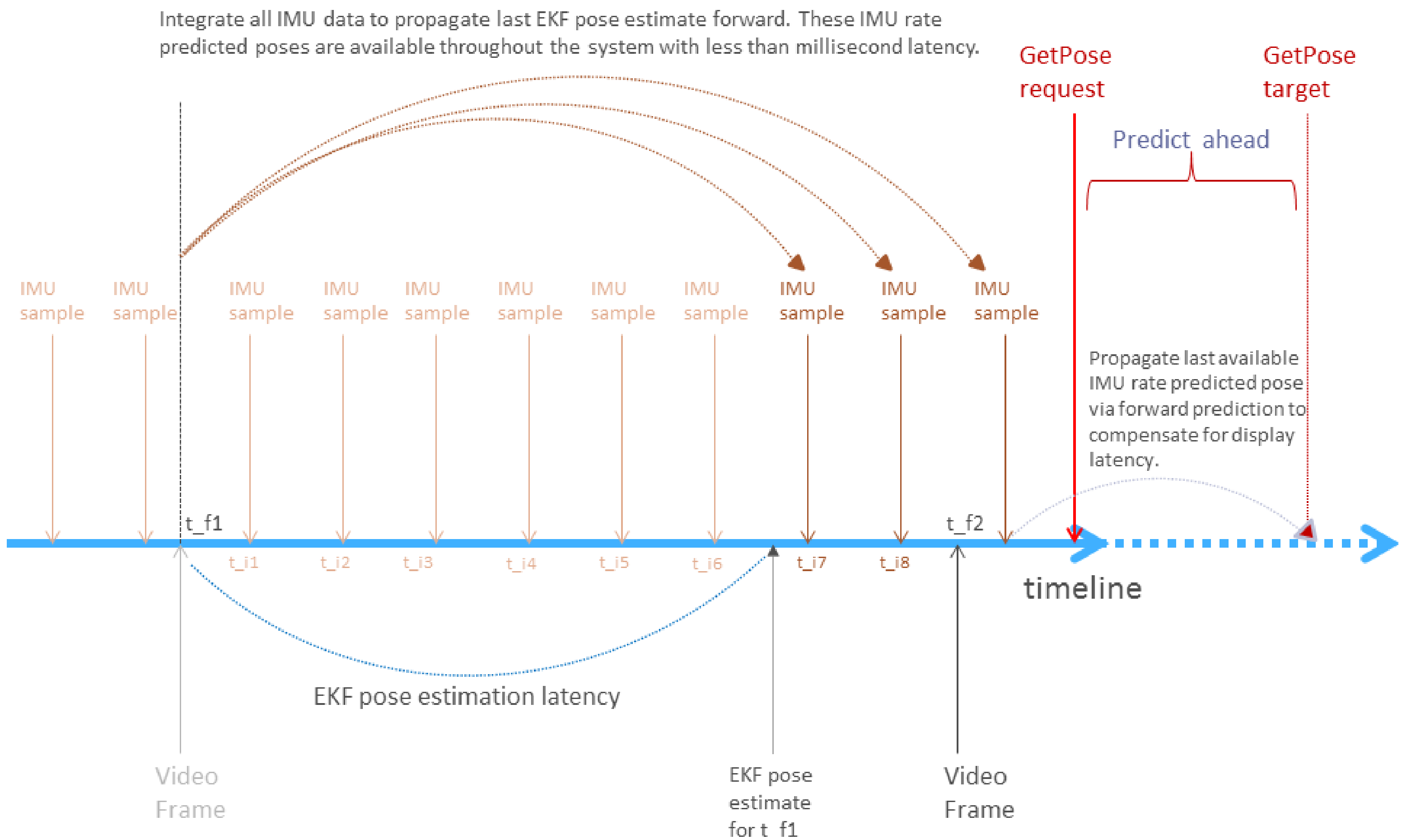

Figure 5. Timeline of events for optical see-through pose prediction

Video based 6-DOF tracking and IMU centric filter. An IMU-centric error-state Extended Kalman Filter (EKF) approach (Kumar 2012) was used to fuse IMU measurements with external sensor measurements that can be local (relative), such as those provided by visual odometry, or global, such as those provided by visual landmark matching, The filter replaces the system dynamics with a motion model derived from the IMU mechanization model which integrates the incoming gyro and accelerometer readings to propagate the system state from a previous frame to the next. The process model follows from the IMU error propagation equations, which evolve smoothly and therefore are more amenable to linearization. This allows for better handling of the uncertainty propagation through the whole system. The measurements to the filter consist of the differences between the inertial navigation solution as obtained by solving the IMU mechanization equations and the external source data. At each update, the EKF estimated errors are fed back to the mechanization module to not only compensate for the drift that would otherwise occur in unaided IMU but also to correct the initial conditions for data integration in the mechanization module. Figure 5 shows the core components that make up the localization system.

The system uses the vision algorithms for both relative pose computation and absolute pose computation. These are both done as inputs in terms of feature based image correspondences to the Kalman filter. The EKF framework uses both relative measurements in a local 3D coordinate system via visual feature tracks and absolute measurements via 3D-2D landmark tie-points as inputs. A 6 DOF pose is computed (both 3D rotation and 3D translation). The visual feature track measurements are applied in a strictly relative sense and constrain the camera 6-DOF poses between frames. Each feature track is used separately to obtain its 3D position in a local coordinate system and a measurement model whose residual is based on its re-projection error in the current frame is used to establish 3D-2D relative constraints on the pose estimate. The 3D location for each tracked point is estimated using all frames in which it was previously observed and tracked. On the other hand, 3D-2D measurements arising from landmark matching are fed to the filter directly and used in an absolute sense for global geo-spatial constraints. Within this framework, the navigation filter can handle both local and global constraints from vision in a tightly coupled manner.

Landmark matching module and landmark database. The landmark matching module correlates what the trainee is seeing with a pre-created visual landmark database to locate the trainee with respect to the Bradley vehicle.

The landmark matching module is divided into two sub-modules: landmark database creation and online matching to the pre-created landmark database. During landmark database creation, a set of video sequences were collected using stereo sensors. From the collected video sequences, an individual landmark database was created for different key locales on the Bradley vehicle. Each individual landmark was characterized by a unique locale ID and object state ID. Locales includes the turret, cargo bay etc. The state ID's included detections such as hatch open/close, tow-launcher shield removed. This database was used for all training events.

During the online maintenance task, landmarks are extracted from the live video for match to the pre-built landmark database. If a match is found, the locale ID is returned. Given the returned locale ID information, the pre-built landmark database can be further constrained or narrowed for the next input query images to obtain both faster and more accurate positioning of the trainee and states.

Low latency prediction module. For optical see through (OST) augmented reality displays, accuracy of the pose estimates alone is not sufficient for an acceptable user experience. The rendered markers also need to appear with very little delay on the display. This is due to the fact that, in the OST framework, the user sees the real work as it is (not an image of it) and hence the equivalent “frame-rate” is essentially very high with there being no-delay in visual perception of the real world. Therefore, the associated rendered markers must satisfy this stringent requirement in order for them to appear jitter and drift free when they are displayed. Otherwise as the user’s head is bobbing, the markers will appear to bounce around in the display since they will be lagging in time. Figure 5 shows the timeline of sensor inputs and algorithm outputs in relation to forward prediction for camera pose estimation. Video frames in general arrive (15 Hz in our case) at a much slower rate than the IMU samples (120 Hz in our case.) The pose estimate that incorporates information from a video frame is in general available after 40-50 msec processing delay. The pose requests from the renderer arrive asynchronously at the highest rate the renderer can accommodate. After renderer receives a pose it is displayed on the see through display after a certain amount of delay which is affected by both the display hardware latency and lag caused due to inefficiencies in the rendering pipeline and video graphic card. In order to compensate for all the latencies in the system, a forward prediction mechanism that estimates the camera pose corresponding to a certain timestamp into the future, given all the information that is available up until the render request, is utilized. For this purpose, forward prediction performs a second-order extrapolation of the orientation using a window of past camera poses.

3D estimation. The use of 3 dimensional space mapping to improve object detections and overlay localizations was evaluated. If stereo cameras are utilized, depth maps from the perspective of the user can be computed. These depth maps can be used to better detect objects and to render overlays taking dynamic occlusions into consideration. The stereo depth computation module uses pyramid based processing to obtain depth maps efficiently.

Scene/event understanding. The scene and event understanding module provides the system with the following capabilities: (1) recognizing some basic maintenance tools and Bradley parts; (2) recognizing some basic states of Bradley parts. The basic maintenance tools recognized are turret drive level, torque wrench (1/2 inch drive, 0-170 ft-lb), and 3/8 inch drive 14 mm socket.

A fast tool detector was trained to detect and recognize the maintenance tools using AdaBoost (as implemented by Rojas, 2009). AdaBoost is an aggressive learning algorithm that produces a strong classifier by choosing visual features in a family of simpler classifiers and combining them linearly.

In addition, the system is able to recognize the Bradley parts from the locale ID matched from the pre-built Bradley landmark database. With the continuously tracked camera pose information relative to each Bradley locale, the system is able to remap back to the 2D image and segment each Bradley part precisely and recognize it. Similarly, since the landmark database for each Bradley part was pre-built at different states, for example with the cargo hatch closed and with the cargo hatch open, the state information can also be easily extracted and returned to the system.

VPA proxy. The VPA proxy module wraps the communications between the SPS and VPA subsystem as a DSF filter. It allows the SPS to receive and send data to the VPA module. It also allows for these messages to be routed to other filter modules. These include messages to the render-proxy and scene/event understanding filters.

Render proxy. The render proxy module wraps the communications between the SPS and rendering subsystem as a DSF filter. It allows SPS to send low latency messages of the trainee position to the renderer. Additionally, messages from VPA are directed from the VPA proxy filter to the render-proxy to be sent to the rendering subsystem.

VPA Subsystem

VPA is a real-time dialog system that supports human-like interaction using spoken natural language. The VPA system recognizes the user's goals and provides feedback to the user. The feedback and interaction occur both verbally and by engaging the augmented reality system to display icons and text visually on the user's eye glasses. The major blocks of the VPA subsystem are shown in Figure 6 below. A knowledge-base was constructed for the particular set of training tasks. The base includes 3D object/ action models for scene understanding, grammar and language models for natural language understanding, task workflows for reasoning, and templates for outputting speech and animations to the user's head worn AR Mentor system. This section provides further details of the design for the knowledge acquisition, natural language understanding, and reasoning modules.

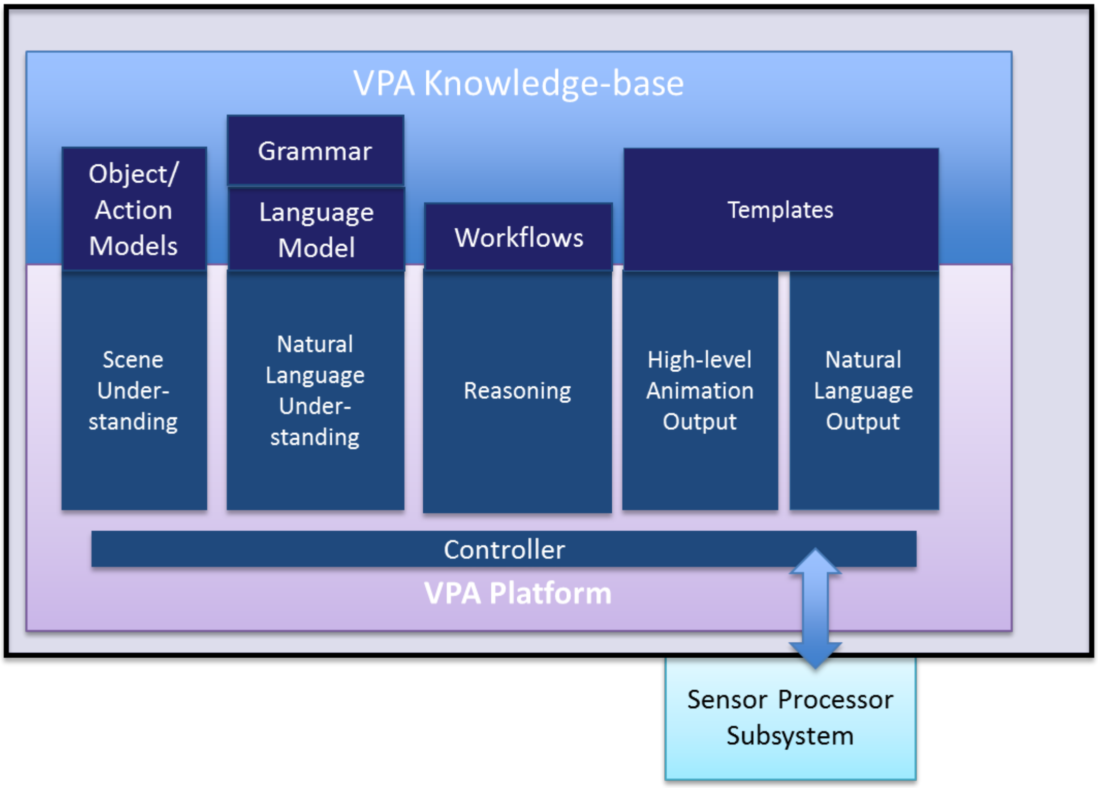

Figure 6. VPA subsystem block diagram

Knowledge acquisition. Multiple sessions were conducted with Bradley maintenance subject matter experts (SMEs) to develop content for the knowledge acquisition module. Information was collected to support the system's interaction with the following features:

- Task model – The system displays each step of the task on the eye glass, and the trainee can reads the step aloud to perform it.
- Verbosity level – The trainee can dynamically switch the verbosity level between detailed instruction and minimal instruction.
- FAQ – The system supports trainees' frequently asked questions
- Show additional information – The system can show the location of a part via an overlay, explain the task step, show the video about executing the step, clarify the purpose of the step, etc.
- Warning and notes – The system can alert the trainee with warnings and notes as the trainee performs the step.
- Intents – The system can impute trainee intents to provide appropriate training contexts. The intent associates objects and actions in a meaningful way in a target domain space. VPA includes several intents such as SelectWorkPackage for gathering the work package information from the user either by the work package number or name, VerifyTool for checking whether

the trainee has all the tools, VerifyEquipmentCondition to ensure the equipment status before starting the training etc.

Natural language understanding (NLU). NLU uses a hybrid understanding strategy for determining the intent as well as associated parameters. It employs:

- High-accuracy, domain-specific rule-based understanding system based on top-down recursive transition network chart parsing, and
- More generic state-of-the-art statistical intent classification and argument extraction systems based on maximum entropy classification.

Rule-based understanding was built using the Phoenix parser (Ward 1994, Phoenix). It uses grammar rules that analyze the order of words and synonyms to determine intent and ignores the words that do not match the rule set. The rule-based parser may return multiple intents. Initially, the grammar rules were developed by leveraging Soldier and SME inquiries for the task during the knowledge acquisition and the role playing sessions. Throughout the project, these grammar rules were enhanced as new data were acquired.

For development of a statistical model for understanding, sample Soldier and SME utterances were collected and annotated with the appropriate intent and arguments. These data were used by the machine learning toolkit MALLET (McCallum 2002) to develop a statistical parser to identify the intent and locate the arguments for a given utterance. Additional data collection during the training session increased the coverage and accuracy.

An interpreter merges the intent extracted from the most recent utterance with the overall intent to create the current user goal within context. AR Mentor tracks previous goals within the current context and uses that information to understand utterances without an explicit intent specified in them. In addition to the default merging, the interpreter workflow can also be customized if necessary.

As an example, if the current intent is VerifyToolCheck and the user says "Torque wrench," the utterance by itself in isolation is not meaningful for the system. However, the interpreter will associate "torque wrench" to the VerifyToolCheck intent.reasoning module.

The VPA reasoning module directs the VPA's actions. In the context of AR Mentor, it guides the trainee through the task he is learning. Task performance information captured in the knowledge acquisition sessions was used to build dialog models. In general, dialog models embody the directions that a prototypical conversation for a given intent can take. In AR Mentor, the trainee's primary intent is to perform the training task, and the dialog model for this intent embodies how the system will converse with the user (including AR interactions) during the performance of that task.

As an example, the TOW lift upper limit switch adjustment task consists primarily of a strict sequence of steps and sub steps, with a single branch in the middle conditional on the results of a task. Accordingly, the dialog model for the task mirrors this structure, with the system guiding the trainee's reading of the manual text and performance of each step and sub step in the task. Figure 7 shows the high-level outline for this task's dialog model. In addition to this primary task dialog model, we defined dialog models for a number of other intents—those representing the various questions that the trainee was to ask during the training on the task.

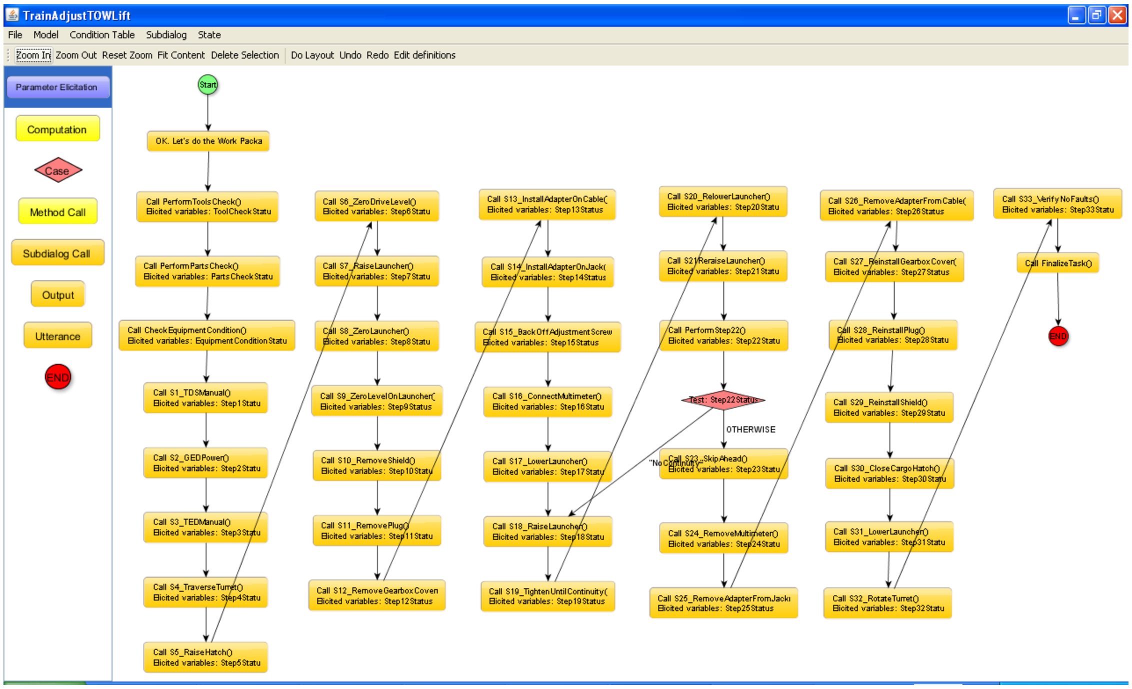

Figure 7. VPA reasoner integrated development environment.

High-level outline of the Dialog Model for the Adjust TOW Lift Upper Limit Switch task. The Dialog Model reflects the task's mostly strict sequence of steps, with a single branch (toward the middle of the task) to repeat a set of previous steps in the case of not observing electrical continuity.

The reasoning module has a number of characteristics that make it amenable to rapid and robust development of new training tasks once the basic AR Mentor architecture is in place. In particular:

- An execution model based on *conditional re-execution*, where the engine performs the bookkeeping necessary to keep track of the results of steps already performed and reason about whether a given task it encounters in the model can be skipped. This makes the Reasoner especially well-suited to training on (or performance of) more complex diagnosis tasks, where the task can take a wide variety of directions, with a wide variety of step combinations, based on the results obtained by steps performed earlier in the task.
- A *reactive component* to the Dialog Models and their execution that allows reaction and response to conditions that can arise at any point during the task.
- Built-in support for easy specification of *context-sensitive responses* to a set of common user questions—e.g., "What are my options?" and "Why do you ask?"--

and other common occurrences—e.g., an utterance from the user that the NLU module can't understand.

Rendering Subsystem

The AR Mentor rendering subsystem is responsible for presenting to the user audio (artificial voice and sounds) and AR rendering visuals (in video see-through or optical see-through modes) to a HMD. These visuals can be of the form of 2D visual overlays, as well as 3D real world anchored indicators. These visuals are meant to present location indicators and directions to a user performing a maintenance task. The presentation system is capable of showing a variety of entity types such as: anchored 3D label, anchored 3D animated models, 2D images, 2D video, 2D text, etc.

The Unity3D[1] rendering system was used as a base layer for the rendering subsystem. Unity3D runs as an independent process, communicating with other processes via plugins. For AR Mentor, a plugin was used to supply the rendering system with a live stream of camera pose data. With this live information, virtual 3D objects could be processed to appear to a user to be anchored to real world positions.

All entity presentation and parameters as well as TTS (text to speech) requests are controlled by the reasoning system and communicated to the renderer via a custom network protocol. This protocol enables the specification of sequences of presentation actions as well as synchronization of presentation actions, enabling the reasoning engine to provide coordinated presentation action timing at the renderer level, where such timing would be best controlled.

A flexible AR command protocol and system is used to control the renderer. This command system allows for a command to be composed of a sequence of AR actions. This sequence of actions acts as a script controlling the presentation and modification of 2D and 3D AR elements. Optionally, execution of actions can be queued, causing subsequent actions to wait on completion of previous actions. This facilitates a method of temporal sequencing of action execution. AR actions include:

- Adding an AR element (Label3D, Model3D, Video 2D, Image2D, etc.), with all parameters, to the scene.
- Removing an AR element or group of AR elements
- Using TTS to generate and play artificial speech
- Waiting for a short period of time
- Directing the user towards an object
- Modifying a parameter (e.g. text, text color, model orientation, etc.)

[1] Unity3D is a trademark of Unity Technologies, San Francisco, CA.

- Calling a function (e.g. StartVideo, PlayModelAnimation) of an existing in-scene AR element. The modification is either immediate or (for certain parameters) can span a period of time, using a provided ease function

Command scripts can be used to encapsulate involved sequences of events. In order to provide a higher level of abstraction to the reasoning engine, a database of command scripts can be loaded and cached at start time or run-time. A special call can be sent to the renderer to invoke these cached command scripts by name.

Unity3D typically requires all assets (videos, images, models, animations, fonts, etc.) to be baked into the application build. It does this to hide the assets from use by other applications and to transform the assets into a normalized form that can be loaded quickly by the run-time. Thus, adding or modifying an asset requires recompilation. To circumvent this limitation, Unity3D Asset Bundles (typically reserved for loading assets from a patch server over the internet) were used. Asset bundles can be generated containing new assets and/or asset updates and loaded into the system at runtime, adding to the available assets that can be used by commands.

NeoSpeech[2] was used as the text-to-speech (TTS) engine due to its high quality. The TTS plugin provides an abstraction level hiding any details of the particular TTS engine, so that the TTS engine can be replaced with another if needed without any renderer code modification.

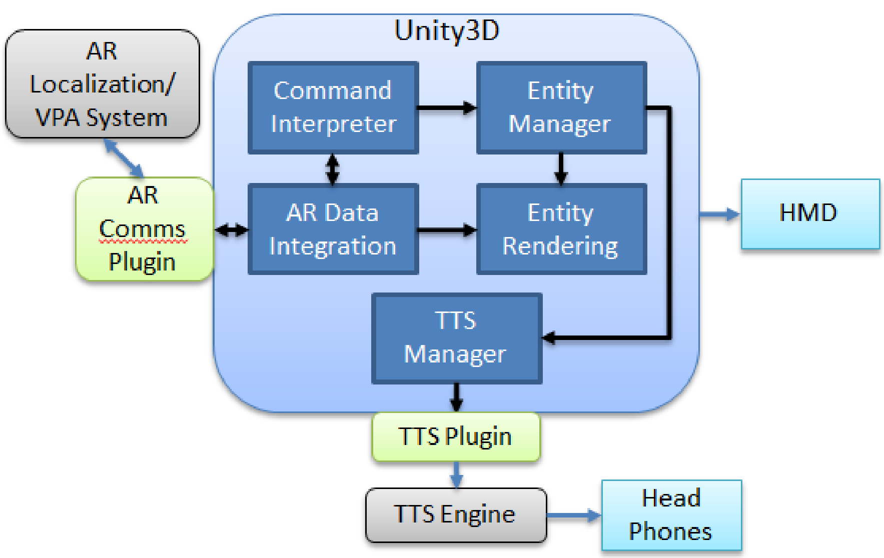

Figure 8. Rendering subsystem diagram

[2] NeoSpeech is a trademark of NeoSpeech Incorporated, Santa Clara, CA.

Interfaces between Subsystems

Communication between VPA and sensor processing subsystems. All communication between VPA and the sensor processing subsystems is conducted via web services implemented by the VPA proxy and VPA server components as shown in Figure 9. This approach allows for the communication to be both cross platform and cross language. The VPA Proxy is implemented in C++ using the POCO C++ libraries. The VPA Server is implemented in Java using the Apache Tomcat web server and jabsorb library. JavaScript Object Notation (JSON) parsing is implemented using the JsonCpp library.

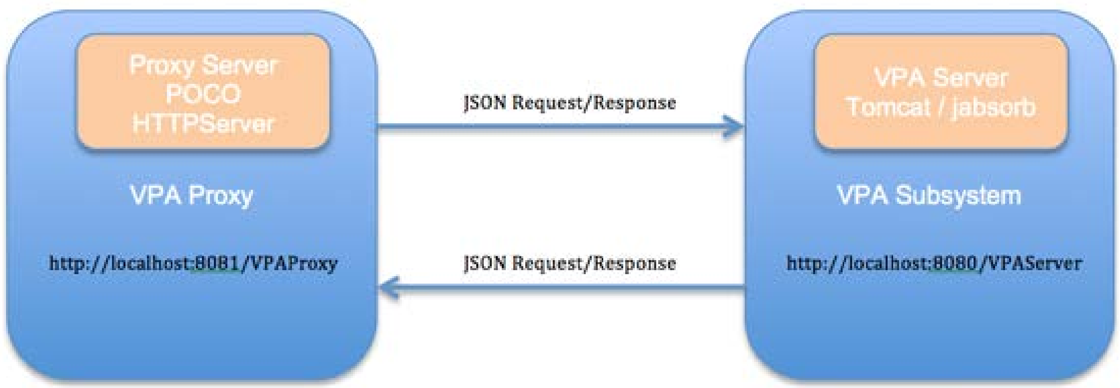

Figure 9. JSON messages used to encode information shared between VPA and sensor subsystems

JSON messages are used to encode the information to be shared between the sensor subsystems and the VPA reasoning components. Examples of these messages might include change of location or gaze events, natural language utterances by the user as well as current task and status updates.

Communication between rendering and sensor processing subsystems. All communication between the sensor processing system and the AR rendering systems (which are separate processes) is handled over a bi-directional shared memory IPC (inter-process communication) first in, first out (FIFO) queue. Shared memory is the fastest available method for communicating among multiple processes, and was used to reduce latency.

The protocol used over this pipe is a custom type-length-value binary protocol. A binary protocol is used to accommodate the amount and type of data that may need to be passed between the applications at high frequency, such as stereo images, depth fields, high frequency poses, etc. Text based protocols would be inappropriately inefficient for such messages. The protocol also allows for string based sub-protocols to be used via the command and status messages. For AR Mentor a custom XML (or JSON) AR entity presentation protocol is used to facilitate visual and audio commands to be forwarded from the VPA reasoner. The protocol status messages are also used in the other direction to return command completion messages that would be forwarded back to the VPA reasoner.

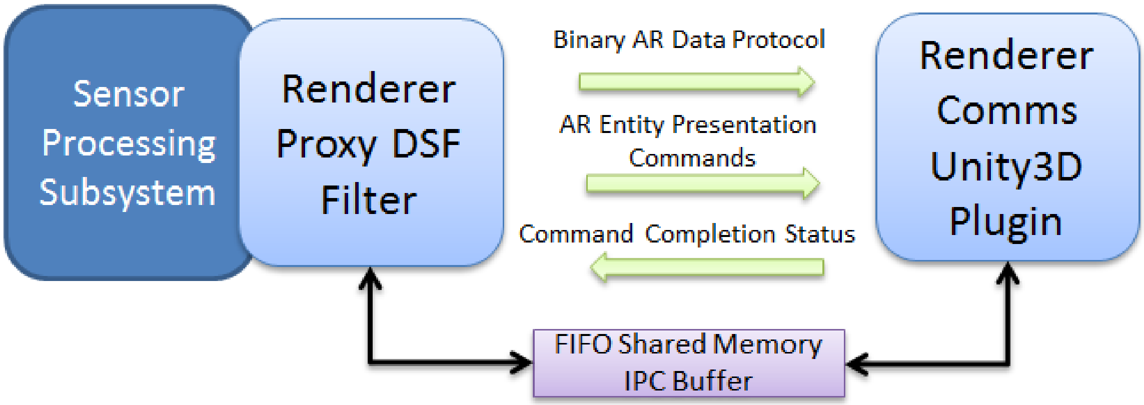

Figure 10. Rendering subsystem communications

Hardware Architecture

The AR Mentor hardware can be categorized into two parts; (a) head mounted sensor package, and (b) AR Mentor processing hardware.

Head mounted sensor package

The primary devices needed in the head mounted sensor package are: a video camera, an inertial measurement unit (IMU), microphone, headphone, and video display. For each piece of hardware, different options were evaluated.

Display selection. Wearable displays are a key element of the system. Two types of displays were evaluated; video see-through (VST) displays and optical see-through (OST) displays.

In VST displays, the real world view is first captured from a camera; the AR overlays are then added onto the captured video; and the fused view sent to a user worn opaque display. The synchronization between the captured real world video and the AR overlays is less of an issue in the video see-through systems. However, limited by current display technology, there is a large latency when displaying high-resolution videos of the real world resulting in perceptible delays in the system's response to user head movements. These delays can hinder the user's task performance.

In OST mode the user sees the real-world through the display while AR overlays are generated and displayed to remain registered to that real-world view as the user moves his head. While OST is the best option, especially for tasks that require the user to observe the surrounding real world with no latency due to the safety concerns (e.g., while working on the Bradley vehicles), any latency in processing the virtual overlays can have significant visible errors. Consequently algorithm burdens are relatively high for the optical see-through system.

In view of the safety demands of the Bradley application, development concentrated on OST hardware solutions. The solution selected was the Cybermind Cyber-I Monocular HMD.

Figure 11 shows the final design of the sensor/HMD packaging built for the optical see-through AR system. It consists of a MicroStrain 3DM-GX3-25 IMU and two Ximea xiQ MQ013MG-E2 (1280x1024 resolution, 63.3^{o} horizontal field of view) cameras as the sensor package. It is integrated with a Cybermind Cyber-I monocular optical see through display unit.**Camera/IMU selection**. The use of one camera vs. two cameras in a stereo configuration was evaluated. Although a stereo camera setup enables better reasoning on the dynamic 3D objects in the trainee's view and allows more complex occlusion reasoning during overlay display, stereo processing is more computationally expensive and requires a larger sensor package than a monocular system. We started with the stereo system to reduce the uncertainty in the first year and then transitioned to a monocular system for the second year.

For the IMU the XSens MTI-G was evaluated against a Microstrain 3DM-GX3-25. The Microstrain IMU was selected because it is able to provide measurement precision equivalent to the XSens MTI-G, but is a much smaller and lighter unit than the XSens MTI-G unit.

Microphone and headphone selection. The microphone and headphones are used by the system to verbally communicate with the user. The user can ask the system questions and hear the answers in the headphones or see the relevant information in the video display. For comfort and ease of wear an over the ear headset was used with noise cancelling technology to reduce background noise.

Secondary Display(s)

Due to current limitations of head mounted displays (low resolution, brightness) and to avoid obstructing the users view more than necessary, one or more tablet systems can be used as HD displays. Typically a secondary display is used to show detailed imagery such as an electrical schematic. The use of a secondary display allows imagery to remain viewable indefinitely without obstructing the user's view (i.e. the user looks at the schematic as needed then looks elsewhere to perform a task.)

A secondary display receives the same set of commands through the VPA proxy as the primary (HMD) display. Since a command set is often initiated by invoking a script name, differences in displayed material was enabled by changing the content contained within a named secondary display script.

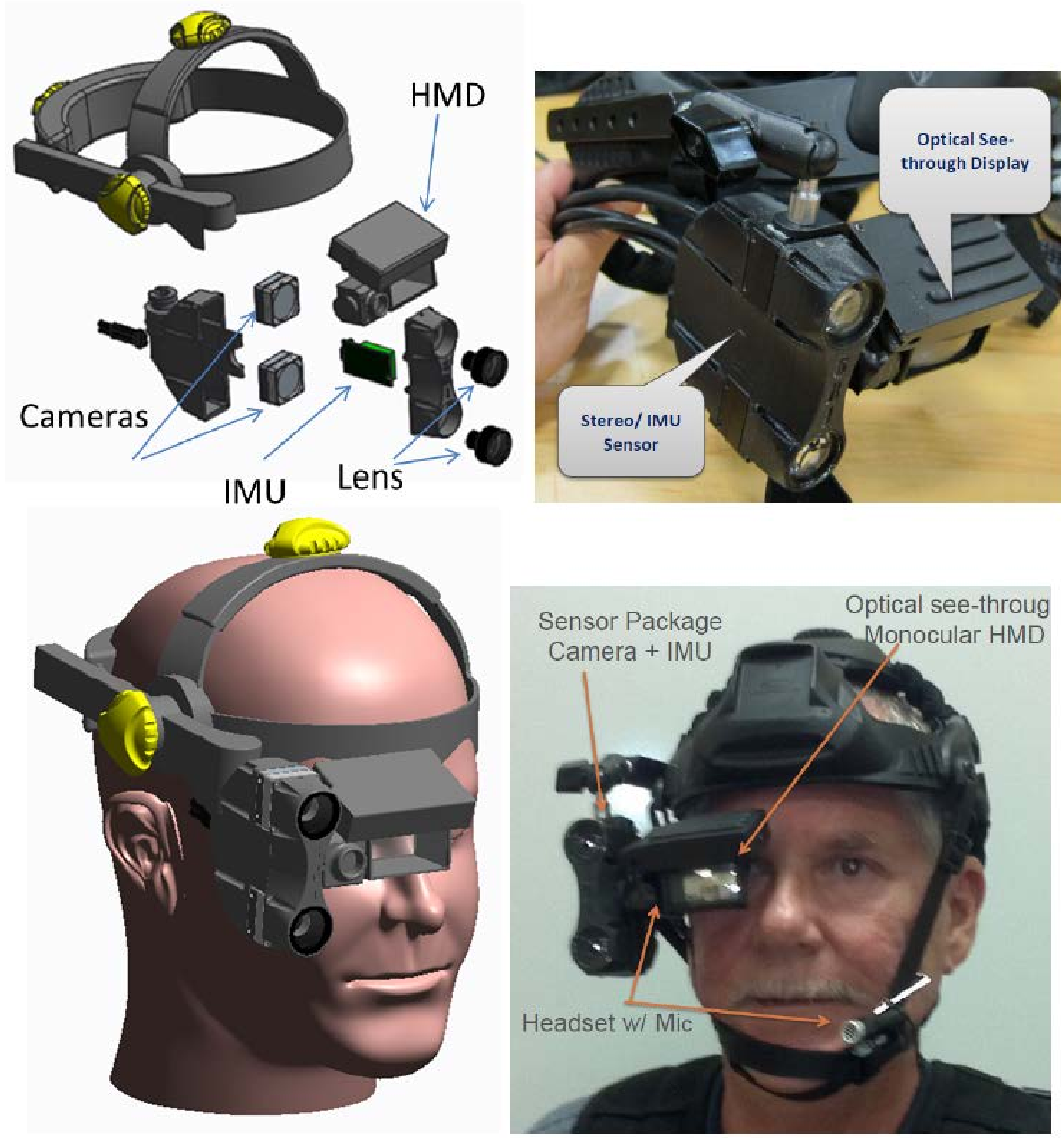

Figure 11. Stereo camera-based design using Cybermind HMD

AR Mentor Processing Hardware

There are three main separate processing requirements which drive the processing capacity requirements for the AR Mentor computing hardware: user tracking and visual processing, natural language processing and reasoning, and rendering. These three tasks require a significant amount of computing power. At the same time there are severe limitations to the size and weight that can be accommodated, given the normal work environment for the maintenance tasks AR Mentor is to address. The approach to developing the processing hardware configuration was iterative across the two years of the AR Mentor project.

For the first year's initial development, a single PCIe-104 form factor Intel i7 quad-core computer was used. The PC along with power filtering boards and HMD interface boards was mounted on the back of a vest to be worn by the trainee. Figure 12 (left) shows the computing package used for the first year demonstration.

Second year development evaluated dividing the processing into a user worn component and a server side component. The user worn component was to provide the video based feature tracking, low-latency filtering, visual landmark matching and AR rendering. The server side system was to provide VPA functionality and ASR. With this division it would be possible to run the user worn component on a mobile processor such as Qualcomm processors used in smart phones (Figure 12, right). However, smart phone hardware available for the second year was unable to properly drive the optical see-through display, so a Mac minicomputer was used instead. Communication between the user worn Mac mini and the server was implemented over Wi-Fi (although use of Bluetooth was also possible). The server system used in the final demo was an off-the-shelf laptop Intel i7 quad-core computer. The server system was placed near by the Bradley vehicle being used within Wi-Fi range of the trainee.

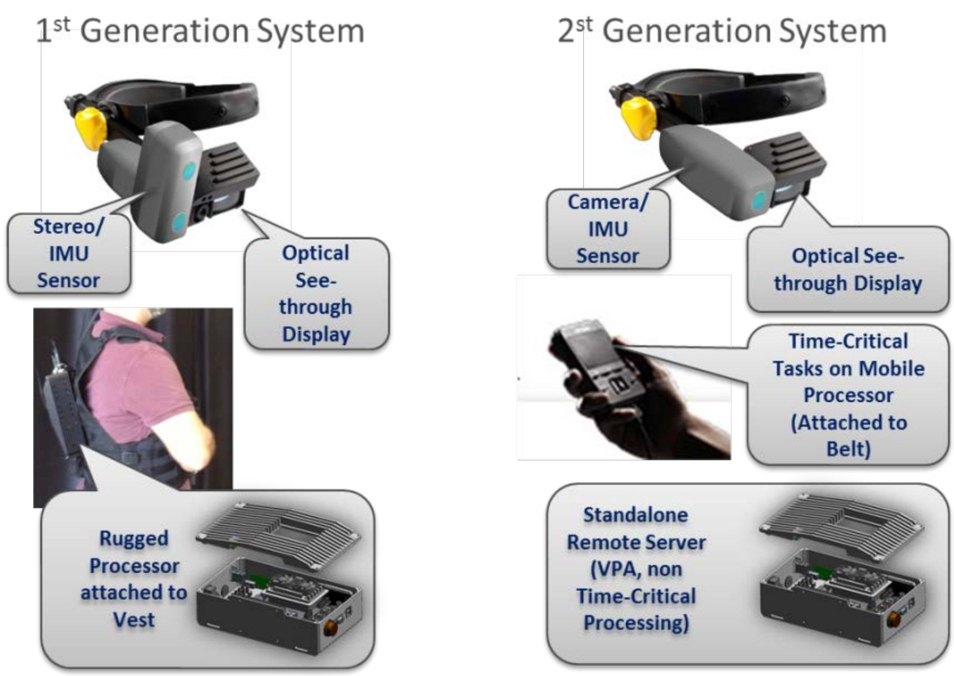

Figure 12. Processing hardware for the AR Mentor system

Assessment for Training

AR Mentor was assessed using students from the Ft. Benning, GA, Bradley Training Division's (BTD) 91M10 course of instruction which trains new Soldier maintainers in basic Bradley maintenance classes. To assess AR Mentor's potential effectiveness for training Bradley maintainers, two evaluations were conducted, one at the end of the first year and another at the end of the second year. Year 1 focused on testing the basic system feasibility for individualized training for detailed maintenance procedures using specific hand tools (e.g., turret level, multimeter), and Year 2 focused on gathering learning outcomes of the system's application to detailed maintenance procedures and on developing a baseline for using AR Mentor to support future use for troubleshooting procedures.

The detailed maintenance procedure selected for training was "adjust TOW lift upper position switch." This 33 step procedure requires both gross movements (e.g. the student moves from position to position on the Bradley) and finer grained activities (e.g. the student connects a test harness and uses a multimeter to check electrical continuity). Students typically train on this procedure in pairs: one student performs the subtasks while the second student reads step-by-step instructions from the maintenance manual. For this task, the AR Mentor VPA modeled the second student; that is, the VPA dialog logic followed the subtask steps as taught in the 91M10 course with no additional pedagogic excursions that might have benefited from potential effects of AR.

The troubleshooting procedures selected for training correspond to a 91M10 block of instruction named "alternate troubleshooting procedures." For this training, students rely on schematic diagrams to isolate/identify faults injected into the Bradley main electrical power distribution subsystem. For troubleshooting procedures, working from a schematic, individual students must be able to locate and identify physical circuits, interconnections, and test points.

Both the maintenance and troubleshooting procedures are taught on specific days during the conduct of the 91M10 course. Students participated in the research at the point in the course at which they would have received training on the procedures during normal course progression.

The Year 1 assessment design included both assessment instrumentation development and preparation for a basic, exploratory feasibility test of learning and performance outcomes. The assessment compared maintainers in a maintenance manual-supported learning condition versus an AR Mentor learning condition. This work occurred in fall 2013.

The Year 2 assessment focused on measuring the learning outcomes achieved in the schoolhouse between maintainers learning the procedures using AR Mentor and maintainers learning in the usual fashion with an instructor. It also focused on gathering basic feasibility data on the use of AR Mentor to teach troubleshooting, a more dynamic task that involves conceptual understanding of the electrical system.

It was expected that the provision of labels and diagrams in AR overlays and interactive dialogs via the VPA would reduce student's time spent looking up and translating information in the technical manual and schematics and would increase time spent on focused “as needed” learning. These features were expected to improve maintainer perceptions of learning and actual performance on learning outcomes of procedural and conceptual knowledge and skills.

There were also some potential negative aspects of using AR Mentor. These included possibly increased training time associated with using technology that focused learners on gaps in their understanding and possibly increased training time due to more enforced practice and iteration.

The work addressed the following questions to provide an understanding of the relative costs and benefits of the two approaches to supporting maintainer performance:

1. What are the relative levels of maintainer help-seeking in the two performance conditions and how successfully can maintainers resolve their questions?
2. What quality of task performance do trainees experience in the two different conditions as measured by (a) attainment of a successful subtask outcome(s), (b) completion of required solution steps per subtask, and (c) time to solution per subtask and across all subtasks?
3. What are trainee perceptions of learning difficulty in the two different conditions?
4. What did AR Mentor participants think of the system in terms of accuracy of diagnostics, timeliness of response, usefulness of response, overall quality of interaction, and what did participants suggest for improvement?

Method

Year 1

A three-group design (AR Mentor only, instructor/technical manual, and technical manual only) was used. The technical manual only condition provided a baseline of key points of students’ learning difficulty without the masking of intrusive instructor guidance. We engaged two groups of participants, 6 novices to show the feasibility of AR Mentor for schoolhouse implementation and 2 experienced mechanics to show feasibility for field implementation. For the experienced mechanics, there was only the AR Mentor condition. All participants conducted the 33-step armored vehicle maintenance task “adjust TOW lift upper position switch” that typically takes a mechanic 40 minutes to perform. In AR Mentor and TM conditions, instructors were asked to avoid intervening; In the Instructor condition, instructors provided guidance as normal. To provide more data in the AR Mentor condition, learners switched roles and repeated the procedure to provide more input on the usability of the system, while pairs went through the procedure only once in the other two conditions.

Figure 17 shows a set of selected example images with AR insertions for maintenance task step 6, which consists of 5 sub-steps that instruct the student how to zero a turret level on a mounting plate step by step. Figure 18 shows a set of selected example images with AR insertions for step 10, which instruct the student how to use a ratchet wrench to remove a shield from a housing by removing four screws in a specific order. Note that these images are frames "grabbed" from a demonstration in video mode.

Figure 13. Example images of Step 6 with virtual insertions (tools, parts and text)

Figure 14. Example images of Step 10 with virtual insertions (tools, parts and text)

Year 2

The second year assessment repeated parts of the first year's assessment of maintenance procedure training, and also addressed AR Mentor's potential for training troubleshooting procedures.

Detailed maintenance procedure performance assessment. A two-group design (AR Mentor only, instructor/technical manual) was used. Twenty-four novice Soldiers assigned to take their regular lesson in the schoolhouse on the maintenance topic were assigned, 12 to six two-person teams in the AR Mentor condition and 12 to six two-person teams with an instructor. Soldiers were balanced between conditions based on a baseline multi-aptitude test; within a two-person team, one Soldier acted as maintainer and the other Soldier as assistant. They had no prior experience with the maintenance topics. Fifteen of the participants were able to switch team roles and repeat least part of the procedure, while the remaining participants did not have sufficient time to switch roles and begin repeating the procedure.

Alternate troubleshooting for novice mechanics. A two-group design (AR Mentor only, instructor/technical manual) was used. Six novice Soldiers learned alternate troubleshooting procedures, with two pairs assigned to AR Mentor and one pair to the instructor condition. Prior to the research session, the instructor/technical manual students had inadvertently received a few hours of instruction on troubleshooting and schematics, instruction that the AR Mentor students did not receive. Additionally, due to unforeseen changes in the pairings, the student pair in the instructor/technical manual condition had higher baseline test scores than the pairs in the AR condition. All pairs were assigned to work on 4 bugs in the power distribution system.

One member performed troubleshooting procedures on bugs 1 and 2, while the other assisted, and then they switched roles and the other member performed the troubleshooting procedures on bug 4, an amalgam of bugs 1 and 2. All Soldiers were then each assessed individually performing a troubleshooting task on bug 3. Figure 15 shows a few recorded images from troubleshooting bug 2 on the engine power distribution task, where the Soldier was to flip the master power switch, and then check the schematics etc. to identify the bug.

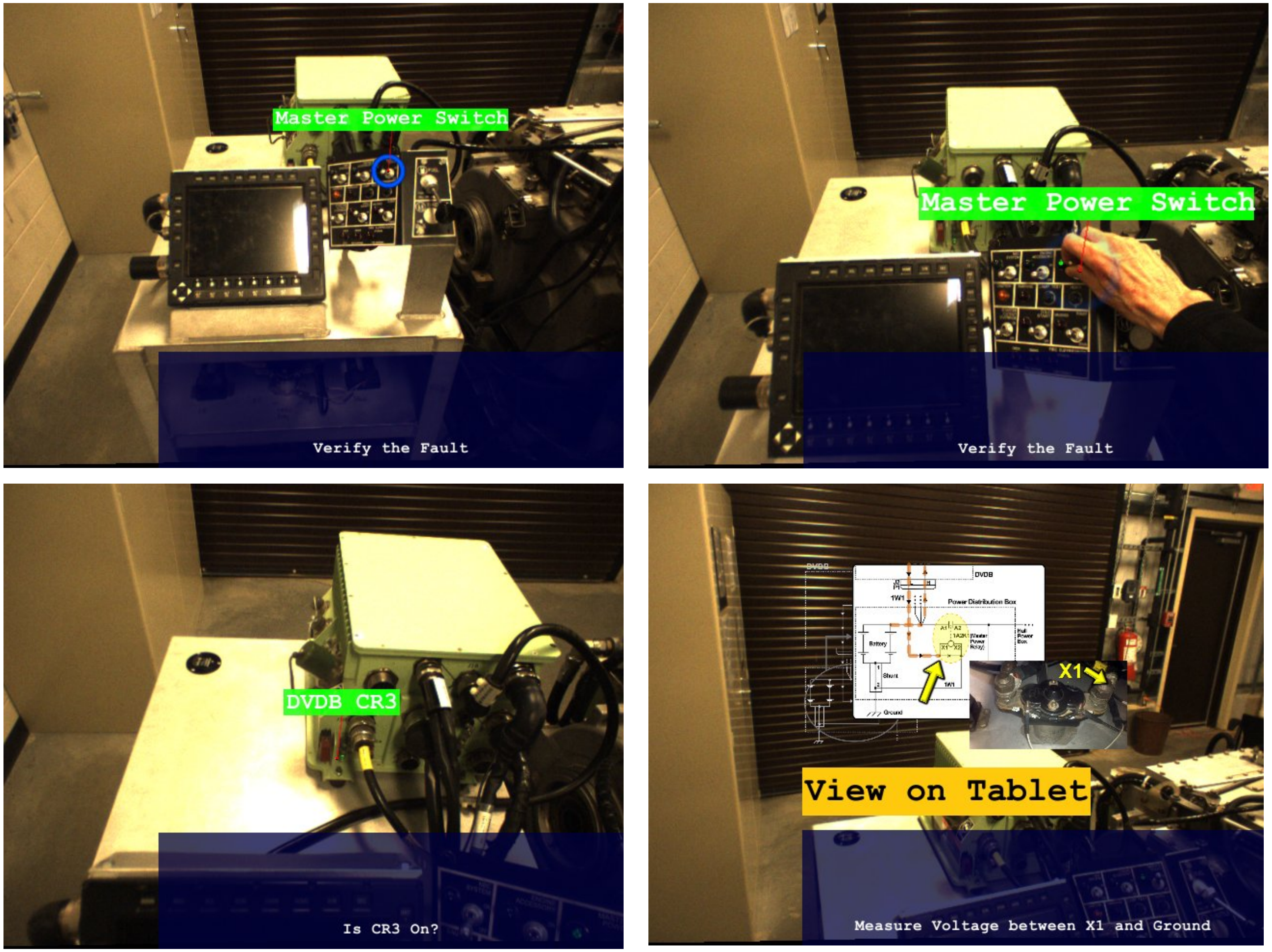

Figure 15. Example image of bug 2 with virtual insertions and instructions.

Assessment Instruments' Administration

Observation protocol and concept checks. Researchers recorded the total time all participants in all conditions took to complete three sub-phases of the detailed maintenance procedure and each of the alternate troubleshooting bugs. Also tallied were the total task completion time, number of errors, and number of instances of help (either sought or intrusively provided). For the instructor conditions, the type of instructor guidance was coded (procedural, conceptual, self-regulating, safety precaution, technical manual correction) and tallied. Researchers interrupted after each of the sub-phases of the detailed maintenance task and posed concept queries (1-3); for the Year 2 alternate troubleshooting case, two interrupted concept checks were done only for Bug 3, which was administered as an assessment to each participant individually.

Learning experience questionnaire. All participants in all conditions filled out a 7-item learning experience questionnaire with a holistic rating scale (1 low – 5 high) of perceived difficulty of the learning experience using a modified, unweighted version of the NASA Task Load Index (TLX) that focused on perceptions of mental demand, physical demand, pace, success of result, effort, frustration, and question-posing difficulty (Hart & Staveland 1988).

AR Mentor usability questionnaire and interview. Those in the AR Mentor condition filled out 19 5-level Likert-scale items asking for ratings of the ease of using the technology's visual and speech features, and answered 3 questions addressing their media representational preferences.

Learning assessment. For the Year 2 detailed maintenance procedure case only, maintainers individually completed a paper and pencil assessment that had 2 procedural sequencing tasks, 6 multiple-choice items, 2 component part identification checkbox items, 1 agree-disagree item, and 3 short response items. A parallel assessment was administered to participants who were available a week later to assess learning persistence.

Results

Analysis followed descriptive and quasi-quantitative methods (Stake, 1995). For both years, researchers tallied behavioral data and compared across conditions and reviewed concept checks. They reviewed mean ratings of learning experience. Based on lags in timing and concept check performance around three sub-steps in Year 1 for the detailed maintenance procedure, a secondary analysis focused on AR Mentor knowledge representations and dialogue density was conducted. Then the engineering team refined the AR Mentor knowledge representations and dialogue pacing accordingly for Year 2. Researchers summarized mean usability scores and interview data for AR Mentor condition for both years, and compared changes from Year 1 and Year 2 for the detailed maintenance procedure case only. For the Year 2 learning assessment, item-level and whole test mean scores were compared between conditions. Classical Test Theory was used to generate p-values for test items. The p-value for an item indicates the proportion of students that responded correctly to the item. Findings are presented in order of the original research questions posed for the performance assessment.

1. *What are the relative levels of maintainer help-seeking in the two performance conditions and how successfully can maintainers resolve their questions?*

As may be seen in Table 1, comparable levels of help-seeking were observed over both years of performance assessment for the AR Mentor as compared to the instructor condition. The Year 1 contrast with the manual-only condition provides a baseline. In both the AR Mentor and instructor modes, maintainers were observed obtaining answers to their questions.

Table 1
Comparison of Total Novice Help-Seeking per Learning Conditions Year 1 and 2 Detailed Maintenance Procedure (DMP) and Alternate Troubleshooting (AT)

Learning Condition	Total Help Seeking Mean Year 1 DMP	*n*	Total Help Seeking Mean Year 2 DMP	*n*	Mean Help Seeking Per Bug Year 2 AT	*n*
AR Mentor	7.5	4	5.63	8	1.83	4
Instructor+Manual	8	2	5.86	7	1.25	2
Manual only	25	2	NA	NA	NA	NA

2. *What quality of task performance do trainees experience in the two different conditions as measured by (a) attainment of a successful subtask outcome(s), (b) completion of required solution steps per subtask, and (c) time to solution per subtask and across all subtasks?*

As may be seen in Tables 2, 3, and 4, trainees made comparable numbers of errors in the AR Mentor condition as compared to the instructor condition but required substantially less instructor guidance. AR Mentor did require a modest increase in time on task. Tables 2-5 present the results both by total and by subtasks of the standard adjustment procedure: Subtask 1 (ST 1) Disassembly and Calibration; Subtask 2 (ST 2) Adjustment; and Subtask 3 (ST 3) Re-assembly. Table 6 presents results of the alternative troubleshooting task. Alternative troubleshooting could not be divided into subtasks as it unfolded as a series of similar decisions to select points to test a single circuit to identify the cause of a fault.

Table 2.
Year 1 Detailed Maintenance Procedure: Errors and time to learn by Subtasks and Total Task

Learning Condition	ST 1 Mean Errors (Time)	ST 2 Mean Errors (Time)	ST 3 Mean Errors (Time)	Total Task Mean Total Errors (Mean Total Time)☐
AR Mentor (n = 4)	2.5 (37:45)	0.25 (11:30)	0 (23:45)	2.75 (1:13:00)
Instructor+ Manual (n = 2)	1.5 (29:00)	0 (10:30)	0 (16:00)	1.50 (0:55:30)
Manual only (n = 2)	4.5 (1:34:00)	1.5 (25:30)	1 (27:30)	7 (2:27:00)

Table 3.
Year 1 Detailed Maintenance Procedures: Instances of Instructor-provided Guidance for Each Learning Condition

Learning Condition	ST 1 Mean Instr. Guidance	ST 2 Mean Instr. Guidance	ST3 Mean Instr. Guidance	Total Instructor Guidance
AR Mentor (n = 4)	1.5	0	0.5	2.00
Instructor+ Manual (n = 2)	28	11	7.5	46.50
Manual only (n = 2)	15.5	6.5	0.5	22.50

Table 4.
Year 2 Detailed Maintenance Procedures: Total Errors and Time to Complete for Learning Conditions

Learning Condition	ST 1 Errors (Time)*	ST 2 Errors (Time)	ST 3 Errors (Time)	Mean Total Errors (Time)
AR Mentor (*n* = 8)	0.64 (0:36)	1.00 (0:15)	0.0 (0:18)	1.75 (1:13)
Instructor+Manual (*n* = 7)	1.33 (0:37)	0.78 (0:18)	0.14 (0:17)	2.00 (1:14)

*Time in *H:MM* format.
.

Table 5.
Yr 2 Detailed Maintenance Procedures: Instances of Instructor Guidance during Training

Learning Condition	ST 1 Mean Instructor Guidance	ST 2 Mean Instructor Guidance	ST3 Mean Instructor Guidance	Mean Total
AR Mentor (n = 8)	1.09	0.63	0.0	1.75
Instructor+Manual (n = 7)	9.78	4.33	1.57	14.71

Table 6.

Troubleshooting: Errors, Instructor Guidance and Time to Complete per Bug

Learning Condition	Mean Errors Per Bug	Mean Instructor Guidance Per Bug	Mean Time Per Bug
AR Mentor (*n* = 4)	1.75	0.63	0:19:00
Instructor (*n* = 2)	0.63	12.38	0:14:00

Year 2 learning assessment results revealed no difference between the AR Mentor and instructor conditions. Test results immediately after the initial training showed AR Mentor condition Soldiers and Instructor condition Soldiers appeared to perform equivalently (AR M = 9.77 out of a possible maximum test score of 14; Instructor M = 9.88, Total n = 23). A test administered a week later indicated a slight decline in Soldier recall, but both groups again performed with statistical equivalence (AR M = 8.77; Instructor M = 8.27, Total n = 20).

Tests of item difficulty using classical test theory, coupled with reviews of the item frequency distributions, indicated that the test overall was relatively easy, with 11 out of 14 items showing that the per item percent correct (p value) exceeded 70%. Test analysis showed that the most difficult items were those focused on recalling the precise bubble on the turret drive level to monitor when leveling a component into a particular position on the vehicle (Item 8 p value = .48 posttest, p value = .40 delayed posttest; Item 9 p value = .70 posttest, p value = .35 delayed posttest) and two short response items asking Soldiers why they calibrated the component with the turret drive level (Item 13 p value = .61 posttest, p value = .65 delayed posttest) and what situation would necessitate an operator request for the adjustment procedure in the field (Item 14 p value = .25 posttest, p value = .20 delayed posttest). The change in difficulty level on Item 9 from posttest to delayed posttest may be attributed to the change in the item from asking learners during the first test to identify the bubbles on the bubble level for leveling the *top* of the vehicle component and then asking the learners in the delayed posttest to identify the bubbles on the bubble level for leveling the *bottom* of the vehicle component. The relatively lower quantitative results on the two short-response items tentatively indicate that the concepts in these items are challenging for learners, although a full check for construct irrelevant variance would need to be conducted to be certain. The assessment found that additional effort in both instructional conditions may be required to help Soldiers understand these concepts. Conceptual knowledge performance was low across both conditions.

In the case of troubleshooting training, final assessment results comparing the AR Mentor to instructor condition had to be discounted because of lack of baseline equivalence between the two study conditions. However, to provide an indication of AR Mentor's efficacy, the 4 AR Mentor Soldiers averaged 44% correct on concept checks in the Bug 3 assessment, and displayed adequate recollection of the procedures for using tools and recognition of components. By comparison, the two Soldiers who had one day's training prior to participating in the study practice sessions averaged 100% correct on Bug 3 concept checks.

3. What are trainee perceptions of learning difficulty in the two different conditions?

For the detailed maintenance procedure, TLX results indicated learners perceived comparably moderate difficulty in learning the task in both the AR Mentor condition and the instructor condition in both Years 1 and 2 as shown in Table 7. The one noted difference was that participants (n = 4) in Year 1 in the instructor condition reported the task was easier than participants (n = 12) in the instructor condition in Year 2.

Table 7.
Detailed Maintenance Procedure: Perceived Task Difficulty under Different Training Procedures.

	AR Mentor		Manual + Instructor		Manual Only
	Year 1 (*n = 4*)	Year 2 (*n = 12*)	Year 1 (*n = 4*)	Year 2 (*n = 12*)	(Year 1 only) (n = 4)
Mean overall difficulty level	2.45*	2.39	1.80	2.63	3.32

* Difficulty scale: 1 – 5, 1=low

For the troubleshooting case, the TLX results indicated participants perceived moderate difficulty in the AR Mentor condition (n = 4) and moderately low difficulty in the instructor condition, (n = 2) as shown in Table 8. The results for the instructor condition may have been influenced by the participants having had the benefit of an additional day of instruction.

Table 8.

Troubleshooting Procedures: Students' Perceived Task Difficulty for AR Mentor vs. Normal Instruction

	AR Mentor (*n = 4*)	Manual + Instructor (*n = 2*)
Mean overall difficulty level	2.61*	2.14

* Difficulty scale: 1 – 5, 1=low

4. *What did AR Mentor participants think of the system in terms of accuracy of diagnostics, timeliness of response, usefulness of response, overall quality of interaction, and what did participants suggest for improvement?*

For the detailed maintenance task in Year 1, maintainers gave high ratings to the 6 types of visual representations—video, text, directional arrows, diagrams, armored vehicle map, and 3D animations (Overall M = 4.39 on 1-5 scale) as shown in Table 9. They rated video and text best. In Year 2, they gave similarly high ratings to 5 types of visual representation—video, text, directional arrows, diagrams, and 3D animations (Overall M = 4.35 on 1-5 scale). They rated directional arrows and text best. For the alternate troubleshooting task, the Soldiers gave average ratings overall to the visual images, particularly for the diagrams (Overall M = 3.36 on a 1-5 scale).

Table 9.
Detailed Maintenance and Troubleshooting Procedures: Average Rated Understandability of AR Visual Features.

AR Visual Feature	DMP Year 1 (*n=4*)	DMP Year 2 (n=12)	Trouble Shoot (*n=4*)
Video demos	4.75*	4.42	3.50
Text	4.50	4.50	4.25
Directional arrows	4.33	4.67	4.50
Diagrams	4.25	4.50	2.75
Map images	4.25	3.67	NA
Troubleshooting steps	NA	NA	4.25
3D animations	4.25	4.33	4.25
Overall Average	4.39	4.35	3.36

* Understandability scale: 1 – 5, 1=low

The dialogue quality of the AR Mentor system ratings were as follows: For the detailed maintenance task in Year 1, Soldiers gave moderately low ratings (Overall *M* = 2.58 on a 1-5 scale) as shown in Table 10. In Year 2, they gave somewhat higher average ratings to the dialogue system (Overall *M* = 3.19 on a 1-5 scale). For the alternate troubleshooting task, the Soldiers gave average ratings overall for the pace of the dialogue system's voice pace and understanding (Overall *M* = 3.50 on a 1-5 scale).

Table 10:
Detailed Maintenance and Troubleshooting Procedures: Average Perceived Pacing Issues with AR Mentor Dialog.

Dialog pacing issue	DMP Year 1 (*n* = *4*)	DMP Year 2 (*n* = *12*)	AT (*n* = *4*)
How often did you want to interrupt the AR Mentor when it was saying it didn't understand you?	2.00*	2.50	2.50
How often did you want to speed up the AR Mentor voice when it was speaking?	1.50	2.83	3.50
How often did you want to slow down the AR Mentor voice when it was speaking?	4.25	4.25	4.50
Overall Average	2.58	3.19	3.50

Note: Scale: 1-5, 1=seldom

Discussion

As indicated previously, access to end users for addressing the usability and effectiveness of the system was unfortunately restricted. The resulting relatively small number of empirical observations means that many of the conclusions discussed here are tentative in that they are based on relatively small numbers. Also, it should be kept in mind that this implementation of AR Mentor strictly supported just the methods of instruction that were in use at that time by the Bradley maintainer instructional cadre. That is, the pedagogy then in place directly drove the development of AR Mentor training, and no pedagogical elaboration or deviation that might exploit potential training features specific to the system was employed. Thus, any positive effect on performance that can be ascribed to use of AR Mentor is likely to be a "lower bound" of effectiveness, because that effect was found despite there being no effort to maximize AR Mentor effectiveness.

Training

With these two considerations in mind, the discussion below follows the outline of the four training related issues listed previously.

Help seeking. Trainees sought additional assistance from their instructor at approximately the same frequency regardless of being trained using AR Mentor or using the traditional instructor and manual method, although trainees training using only the maintenance manual sought instruction at a much higher frequency. Although this finding does loosely indicate AR Mentor and traditional instruction result in similar assistance seeking behavior, it should be kept in mind that with these data it was not possible to determine if trainees sought assistance for different topics depending on method of training. Any similar follow-up work should investigate help seeking at a more granular level.

Task performance. For the detailed maintenance procedures, at Year 2, trainees performed approximately at the same level, regardless of training condition (AR Mentor or traditional instructor with manual). The disparity in instances of instructor-provided guidance between AR Mentor and other methods is an artifact of instructor method: for AR Mentor, instructors provided guidance only for safety reasons or when it became obvious a trainee was "lost," while, for traditional training, the instructor provides a running commentary as the trainee completes the task.

However, for both training methods, trainees' knowledge of the maintenance task at the end of training and then again a week later was equivalent. This could indicate that the additional instructor guidance given the traditional group was not needed, or, alternatively, that the AR Mentor training in some manner compensated for instructor guidance.

Perceptions of learning difficulty. Trainees found both the detailed maintenance and the troubleshooting tasks' difficulties to be about the same, regardless of whether training was by AR Mentor or by traditional instructor methods. However, it was not possible to determine if AR Mentor and traditional trainees found the same of different parts of the procedures to be difficult to perform. Any similar follow-up work should address learning difficulty at a sub-task level.

Perceptions of system features. Trainees' rated understandability of AR visual features (e.g., super-imposed text, 3D animations) was high except for the understandability of diagrams for troubleshooting tasks. For troubleshooting tasks, AR diagrams were electrical schematics for the Bradley main power system. Because trainees at this point in their training had only recently been introduced to electrical schematics, it is unclear whether the low-rated understandability was due to the AR representation or due to trainees' general lack of familiarity with schematics.

With regards to voice interaction with AR Mentor, trainees expressed concerns with the pacing of the dialog – at many points they felt that AR Mentor's voice was either too slow or too fast. Also, trainees expressed a mild desire to be able to interrupt AR Mentor's speech.

AR System and Subsystems

With regard to AR functionality, the AR Mentor system appeared to perform acceptably in the sense that none of the trainees reported the AR features as being unacceptably unrealistic or unusable.

Because AR Mentor was developed as a prototype, comments relative to its physical configuration were not solicited. From Figure 11 it can be seen that many of the physical components can be integrated and miniaturized. Also, the components represent the technical state of the art at the time the prototype was developed; any follow-on instantiation would have available to it any technology advances since that time, especially in the area of visual display.

The modular functional software architecture (see Figure 3) lends itself to ease of update, for example, if a voice processing replacement were to be substituted for the DynaSpeak module. However, utilization of the system-specific DSF and AR communications protocol (DSF sample and ARcomm in Figure 3) will restrict the direct portability of the software system.

References

Chaudhri, V., Cheyer, A., Guili, R., B. Jarrold, B., Myers, K., & Niekrasz, J. (2006). *A Case Study in Engineering a Knowledge Base for an Intelligent Personal Assistant.* Paper presented at the Semantic Desktop and Social Semantic Collaboration Workshop (SemDesk'06) at the International Semantic Web Conference. Athens, GA.

Chandler, P. (2009). Dynamic visualizations and hypermedia: Beyond the "Wow" factor. *Computers in Human Behavior*, *25*(2), 389–391.

Dynaspeak [Speech recognition software]. Retrieved from http://www.speechatsri.com/products/dynaspeak.shtml.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139-183.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animation and external animation in understanding mechanical systems. *Cognition and Instruction, 21*, 325–360.

Henderson, S. J., & Feiner, S. (2009). Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR09)*, Orlando, FL, 135–144.

Kim, Y.-D., & Moon, I.-Y. (2013). E-training content delivery networking system for augmented reality car maintenance training application. *International Journal of Multimedia and Ubiquitous Engineering, 8*(2), 69-80.

Kumar, R., Samarasekera, S., Acharya, G., Yarnall, L., Zhu, Z., Wolverton, M., Branzoi, V., Murray, G., Vitovitch, N., Villamil, R., & Carpenter, J. (2014). Augmented reality virtual personal assistant for training, maintenance, and repair. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2014)*, Orlando, FL.

Kumar, R., Samarasekera, S., Chaudry, A., Zhu, Z., Chiu, H.P., Oskiper, T., Villamil, R., Branzoi, V., Hadsell, R., Pursel, E.R., Dean, F., & Garrity, P. (2012). Implementation of an augmented reality system for training dismounted warfighters. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference (I/ITSEC 2012)*, Orlando, FL.

Lee, K. (2012). Augmented reality in education and training. *TechTrends*, *56*, 13–21.

Mayer, R.E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning & Instruction*, *12*(1), 107–119.

McCallum, A.K, (2002). MALLET: A Machine Learning for Language Toolkit [Software tookit]. Retrieved from http://mallet.cs.umass.edu.

Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D., Morley, D., Pfeffer, A., Pollack, M., & Tambe, M. (2007). An intelligent personal assistant for task and time management, *AI Magazine 28*(2), 47-61.

Phoenix Server [Speech parsing software]. Retrieved from http://wiki.speech.cs.cmu.edu/olympus/index.php/Phoenix_Server

R. Rojas, R. (2009). *AdaBoost and the Super Bowl of classifiers: a tutorial introduction to adaptive boosting.* (Technical Report, Freie University, Berlin).

Ward, W. (1994). Extracting information from spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, 83-87

Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., & Preusche. C. (2013). Augmented reality training for assembly and maintenance skills. *Robotics and Autonomous Systems*, *61*(4), 398–403.