Situational Awareness Measuring Method in Simulated Combat – a Case Study

Uriel Huri, Yoav Yulis, Yisachar Shapira
IDF Ground Forces Command Battle Laboratory
Tel Aviv, Israel
urielhuri@gmail.com, yoavyul@gmail.com, Ysa.m.p.z.129@gmail.com

ABSTRACT

Measuring situational awareness (SA) has been studied greatly in recent years; it plays a critical role in decision-making processes, especially in the battlefield, where the commander's knowledge and understanding of the environment is of vital importance to successfully accomplishing the mission. In the digital age in which Command, Control, Communication, Computers, and Intelligence (C4I) systems for improving situational awareness are developed, measuring SA is even more necessary.

The Israel Defense Forces (IDF) Ground Forces Command Battle Laboratory (i.e., the Battle-Lab) is a research-oriented, virtual environment, where active duty soldiers participate in simulated combat on a large scale. This environment facilitates collecting data and controlling issues. In recent years, the Battle-Lab has developed a method for measuring situational awareness with a unique quantitative objective index.

Data collection was carried out on a digital map, in which participants marked situational factors during a warfare scenario. The data addressed their perceived geo-location, friendly forces, and enemies to be. Every measurement was documented with an exact time stamp and location, allowing the data to be compared to the “real situation” in the simulation.

Analysis of the data was made by translating the differences between the participants' marks and the corresponding real situation to a quantitative value. This translation occurred using a deterministic algorithm and a machine learning algorithm based on expert assessments. By employing a simple computer-based process, researchers were able to measure situational awareness and express it as an objective, quantitative rating.

This method of measuring and examining situational awareness will be presented in this paper, along with the results and the analysis of the process applied during an experiment conducted at the Battle-Lab in February of 2018.

ABOUT THE AUTHORS

Uriel Huri is a data analyst. Uriel studied Biology B.A. in the Technion, the Israeli Institute of Technology. Currently, Uriel works as a research assistant in the IDF's Ground Forces Command Battle Laboratory and leads the research team in the Lab. His research includes experiments involving situational awareness and developing research tools and methods at the IDF.

Yoav Yulis is an industrial and management engineer and military systems expert. Yoav holds B.Sc degree from Ariel University. Currently, Yoav serves as senior researcher at the IDF GFC Battle-Lab and leads experiments dedicated to human-machine interactions in complex, multi-operator, military systems.

Yisachar Shapira holds an MBA in Business Administration from Ben Gurion University and a B.Sc in Electricity & Electronics Engineering from Ariel University. Currently, Yisachar serves as the head of the IDF Ground Forces Command Battle-Lab.

Situational Awareness Measuring Method in Simulated Combat – a Case Study

Uriel Huri, Yoav Yulis, Yisachar Shapira
IDF Ground Forces Command Battle Laboratory
Tel Aviv, Israel
urielhuri@gmail.com, yoavyul@gmail.com, Ysa.m.p.z.129@gmail.com

BACKGROUND

The Battle-Lab

In effect, Israel's Ground Forces Command Battle Laboratory (i.e., the Battle-Lab or Lab) is a single-simulator research laboratory. Its purpose is to simulate ground-force combat scenarios into which novel combat systems can be deployed and the impact to scenario progression evaluated. The Lab provides a simulated virtual environment in which new systems can be vetted before being tested out in the field, or even before they are entirely thought out. As a result, concepts can be researched (or dismissed) before a considerable amount of money has been spent to bring it into the physical world.

The Lab's main purpose is to develop, simulate, and study the behavior of soldiers and commanders in military situations using advanced technologies. The Lab relies on an array of game-like combat stations for human participation, supporting as many as 100 concurrent and cooperating participants at one time. These stations are not too dissimilar from modern, first-person perspective video games, but they aim to achieve tactical realism, rather than enjoyable entertainment.

The Battle-Lab examines novel concepts and technologies by conducting simulated, human-in-the-loop experiments and then analyzes how the participants' performance, emotions, and thoughts were affected during the experiment. In addition to its development teams, the Lab maintains a research team responsible for the collection, processing, and analysis of the experiments' data. Using SQL and Python-based tools, the team examines the experiments' data to find answers to the research questions.

A Battle-Lab experiment usually lasts a week or two and involves dozens of participants. Each day of the experiment is divided into different scenarios, where the region of the combat, missions, and the use of the concepts and technologies being examined varies with each scenario. The variety of situations helps to isolate the effects of those concepts and technologies and leads to more dependable conclusions.

Situational Awareness Research – Theoretical Background

Situational awareness (SA) has been studied since the 1980s with initial studies in the field of aircraft, and later widely adopted by human factors scientists. Over the years, SA has become a matter of interest in such diverse areas as air traffic control, nuclear power plant operation (Flin & O'Connor, 2001), vehicle operation, and anesthesiology (Gaba & Howard & Small 1995). Situational awareness is also an important aspect of robotics research, in which autonomous cooperating machines share their knowledge of the environment to constitute a better collective awareness of their surroundings.

Our work focused on the principles in the research of Dr. Mica Endsley, one of the primary investigators of SA. Dr. Endsley defined situational awareness as the **perception** of the elements in the environment within a volume of time and space, the **comprehension** of their meaning, and the **projection** of their status in the near future (Endsley, 1995b). Accordingly, she developed a model representing SA in three levels, which became popular for describing and studying SA.

The first level of the model, known as the **perception** level, is the basic level of situational awareness. It is about perceiving the status, attributes, and dynamics of relevant environmental elements and, therefore, involves the processes of detection and simple recognition.

The second level is the **comprehension** level and requires one to quantify how the perceived information may affect the status and objectives. This level is influenced by the individual's ability to integrate and evaluate information, and the capability to fuse parts of data into an understandable mental picture.

The processes of the first and second levels lead to the last level of SA, **projection**, which predicts the changes of the relevant elements in the operational situation. This level requires the ability to understand how present status and attributes of an element are likely to change in the future.

The Necessity of Measuring Situational Awareness

Situational awareness can be described as one's mental picture, or situation image, created by both memory and understanding of the situation, and which constitutes the basic guidelines in the decision-making process. The general opinion is that as this mental image gets more accurate and detailed, the chances of making good decisions improves.

In the battlefield, commanders are responsible for making decisions that can literally be a matter of life and death. Hence, studying a commander's ability to establish an accurate mental situation image has become imperative for armies all around the world. Measuring and quantifying this ability is necessary to allow comparison between SA of different participants, or in diversified conditions, in order to understand what influences it.

In the information age, wide-scaled intelligence is wished to be brought to commander in the battlefield, to be used for command and control processes. This has driven the rise of using advanced technologies to provide greater access to useful data while in combat. Modern Command, Control, Communication, Computers, and Intelligence (C4I) systems, for example, have become popular as they are thought to play a critical role in creating accurate SA and thus supporting decision making. They facilitate receiving information about rapidly changing operational situations in real time and make massive amounts of data accessible to maneuvering troops and their commanders while in the field.

As a result, a commander is expected to be able to use this information to appreciate the varied elements and understand the dangers that can affect his forces. Thus, it is of great importance to research the human's capability to recognize and comprehend relevant data from these systems. It is also required to test one's learning process when using training tools that target the improvement of the effective use of C4I systems. Those needs demand a reliable and simple index to compare SA in diversified conditions, or when using different methods and tools.

Methods to Measure Situational Awareness

Although many methods were developed and tested over the years, situational awareness measurement is considered to be quite a challenge due to its dependency on multiple variables. Generally, measurement can be classified as one of two primary types: direct measurement and inferred situation awareness methods. While direct measurement tests the outcome of actions reflecting an individual’s SA, other methods approach the processes involved in achieving SA.

The direct measurement approach relies on both objective and subjective methods. Some of these methods are "interfering" methods, in which the simulated scenario is paused until the end of the measurement, and "non-interfering" methods, which are used either during scenario execution, or after it is finished.

Objective methods allow evaluation of one's SA by comparing the "ground truth" and the participants’ knowledge. Participant query methods, such as the Situational Awareness Global Assessment Technique (SAGAT), provide a fixed score to the participant's knowledge of pre-defined environmental elements. SAGAT is an example of objective-interfering query method, since all action is stopped while the participants answer the questions.

Subjective methods of direct measurement, rely on a participant’s own assessment of their situation awareness or observation by experts who rate the individual’s SA according the observer's understanding. A good example is the Situational Awareness Rating Technique (SART), a questionnaire that follows Endsley’s three levels model, and allows the participants to rate their awareness in different categories and appreciate their own SA. It is a "non-interfering" method, because it is used only at the end of the simulated combat.

These popular methods of measurement were found challenging for use in simulated combat for assorted reasons. First of all, most of the direct objective methods require high usage of resources. The SAGAT query, for example, has to be customized to fit each scenario. The customization process takes time and must be done by a person who holds an understanding of the purpose of the exercise and recognition of the information and elements which might influence the ability to achieve this purpose. Moreover, since SAGAT queries must map to each of Endsley’s three levels, its questions have to be formed to address different phases of every relevant element. These requirements increase the complexity of creating and analyzing each query.

Another issue worth mentioning is that SAGAT, along with other interfering methods, forces a pause of the simulated combat, which may cause undesired effect on the scenario's progress. Alternatively, non-interfering methods are usually affected by the subject's ability to recall information about varied elements at different times, rather than his ability to perceive and comprehend information in real time.

There are subjective methods which allow SA evaluation in real time. They do not provide an indication of how well the subject knows his environment, but rather how well is his SA in either **his own opinion** or **his observer's**. Thus, their results are not comparable when examining SA of different participants in varied scenarios.

THE NEW METHOD: QUANTITATIVE INDEX RATING TECHNIQUE

Characterization

The goal of the Quantitative Index Rating Technique (QIRT) was to resolve the difficulties the Lab had faced with the application of other methods. Thus, the new method had to be simpler to use, with the least amount of interference in scenario execution as possible. Additionally, it had to provide an objective index that could be used to compare SA scores of different participants in varied scenarios.

To reach these objectives, we decided to focus the measurements on the first level factors, testing only how well the participants could perceive their surroundings. Consequently, the measurement process was shorter and easier to deal with, both by the participants and the data collection team. To get objective results, we designed and employed algorithms that use more general factors that reflect SA, such as perceived self and friends' locations, and translated them to a quantitative rating.

Development Process

Following the characterization and definition of the goals of the project, development of the QIRT included five (5) steps.

Steps 1 and 2 were about collecting data. Two types of data were needed for further analysis: commanders' SA data in the simulated battlefield, and experts' ratings of the commanders. The commanders' SA data was gathered by a collection team. After cleansing the data, some of the samples were delivered to experts with a rating sheet. The experts' ratings were collected and organized for use. Due to time issues, we could not deliver all of the samples to the experts.

Since our method is pioneering, we used three (3) methods to analyze the experts' ratings data. In the first method, a single subject matter expert (SME) rated all of the samples by examining every mark of the subjects. In the second method, we used the principles of a different subject matter expert to design a deterministic algorithm. The last method used the ratings of the experts by machine learning algorithms in order to construct a reliable model for calculating one's SA. Steps 3 through 5 describe each method in greater detail.

Step 1: Collecting Situational Awareness Data

In February of 2018, the Lab conducted an experiment with about 150 participants. The forces in the experiment were divided into four companies, with three to four platoons each. There were two primary kinds of officers: platoon commanders and company commanders. Each platoon was either infantry or armored. The participants were all males between ages of 20-30 years old and all served as officers in the IDF. Since the research was about commander's SA, we tested only the officers. There were 19 officers in total.

The primary challenge in collecting the data was limiting interference in the scenario. The main problem of many SA measuring methods is that in order to question the participants, the scenario must be paused or finished. This interruption could impact the management of the combat inappropriately, and therefore affect the results.

To mitigate this risk, we dedicated a team exclusively to the data collection part of the experiment. This team consisted of seven (7) members. Four members, the observers, were placed near the heads of the companies and their task was to identify suitable times for testing that would cause the least disruption to the scenario. The remaining team members, the testers, were positioned in accessible places around the room, with a computers on which the SA tests were conducted.

Moreover, we identified situations in which an officer could be tested with the least interference to the combat. These situations included times when forces were conducting a routine action, when no enemy is around the forces, right after contact, or after simulated death. Each of these simulation states represented a time when the subject was not playing a main role in the combat and, therefore, had less influence on the scenario progress.

The testing was performed by an in-house software application available in the Lab. The application is web-based blank map and includes a bank of forces. Each test participant had to place the relevant forces from the bank at their current position on the map. All of the marks are saved to the experiment's SQL database.

Once an appropriate situation had occurred, one of the observers called to test the participating officers. A tester took the relevant officers to the testing positions and conducted the test. The participants placed the red and blue forces on the map and returned to their positions to continue the battle.

Every subject had to mark the locations of his commander, his subordinates, his peers, and himself on the map. In addition, every subject was asked to mark up to five enemies around him.

In the experiment, 122 samples were collected. Due to technical issues, some of the samples were not useable. As a result, only 72 verified samples were used for further analysis.

Step 2: Collecting Experts' Ratings

Initially, we identified individuals with the right qualifications to rate the SA of the subjects. Each expert had extensive experience in combat and commanding. Ultimately, eleven experts were chosen to rate the commanders' SA.

As mentioned earlier, due to time issues we chose to present to the experts 22 out of 72 samples. We also decided to focus on friendly forces locations.

To simplify the rating process, we created a PowerPoint document and a rating sheet. Every page in the presentation contained a title, a forces bank, and two pictures (see Figure 1). The title detailed the sample number and the role of the participant who marked the sample. The forces bank listed the forces that the participant was supposed to mark. One picture was an image of the subject's marks on a map and the other was a photo of the actual simulation.

The expert examined the pictures and compared the officer's marks to the actual locations in the scenario. The expert then rated the SA of the officer. The rating sheet included a column with the sample numbers and two blank columns in which the expert filled in the rating of the sample and short explanation about rationale for the rating (see Figure 2).

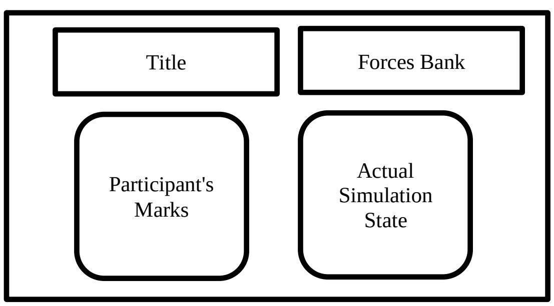

Figure 1. PowerPoint Page Template

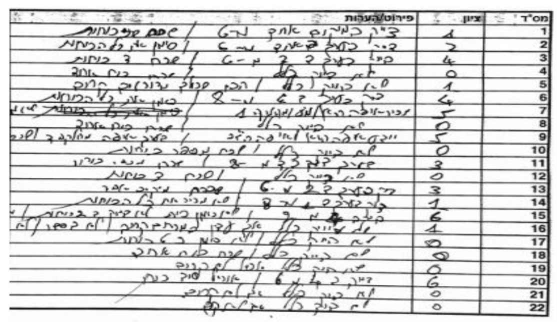

Figure 2. Example Rating Sheet (note that the page is written in Hebrew)
From right to left, first column is the number of the sample, the second is where the expert fills his rate, and the third column is for comments and explanations.

Step 3: Analyzing the Data by a Professional

The first method of analysis is similar to that used in existing SA measuring methods. Unlike the eleven experts who rated the 22 chosen samples, the designated SME reviewed each of the 72 samples in order to supply professional analysis of all of them. In a similar way to the presentation, every sample was rated with an explanation of the rating.

The SME based his ratings on four parameters: role, corps type, scenario state, and terrain type. He also referred to both friendly and enemy forces which were marked by the participants, and separated his ratings between friendly situational image and enemy situational image.

In order to compare each sample to its matching combat situation, we presented the marks and the entities' actual locations on a virtual map tool. Furthermore, the SME used log files and transmission records of the experiment to examine samples carefully.

Finally, the average and standard deviation from different cross-sections were calculated and compared. For example, the average rating of the company commanders was compared with the average rating of the platoons' commanders, and the average rating of the armored forces was compared with the average rating of the infantry commanders.

This process provided the advantage of being thorough. All samples were rated and included a detailed explanation for each rating. Moreover, the results are comparable, since they are all based on the same person's subjective observation. The primary shortcoming is that it took a long time and a larger number of samples would have taken even longer.

Step 4: Analyzing the Data using a Deterministic Algorithm
One of the eleven experts who rated the samples in step 2 specializes in the area of analyzing commanders' SA. He has written guidelines to an algorithm which, in his opinion, represents the quality of a commander's SA. Following his own principles, the expert rated the 22 samples we had given him.

The rate of each sample was determined using six parameters*: Self to Mark, Friends to Mark, Friends Direction, Commander to Self, Commander's Direction,* and *Required Marks Percentage.*

Self to Mark is simply the distance between where the subject marked himself to be and where he actually was.

Friends to Mark is the average difference in distance calculated between subjects and friendly forces. First, two distances between self-force pairs were determined: the distance between the mark of the friendly force and the subject's self-mark, and the actual distance between the location of the friendly force and the self-location in the simulation. The difference of those values was then calculated (see Figure 3). The final value was the average of all the differences in the sample.

Friends' Direction is the average of the error angles in the directions of the friendly forces relating the self-mark of the subject (see Figure 4).

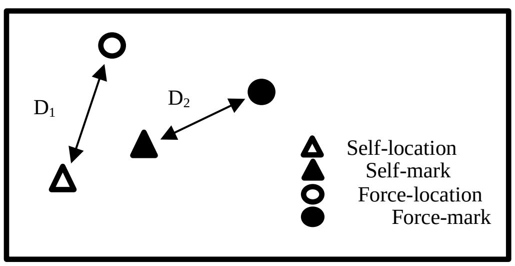

Figure 3. For every friendly force mark, if D_1 is the "real" distance between the force and the subject, and D_2 is the distance between the marks, then the difference is $|D_1$-$D_2|$

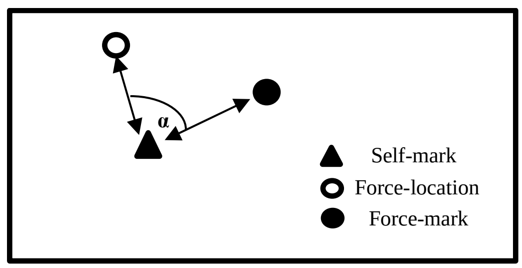

Figure 4. α is the error angle of the direction

Commander to Self is the difference between the actual distance between the subject and his commander in the combat, and the distance between the locations as the subject marked them.

Commander's Direction is the error angle between the commander's actual location and his marked location, relative to the subject's self-mark, as presented above for *Friends' Direction* (see Figure 4).

Required Marks Percentage is the relation between the number of marks the subject actually used and the number of required marks.

In order to get a final score, each of the parameters was considered with adjusted weight and a grading scale. In order to calculate both of these values for each parameter, we ran the algorithm hundreds of thousands of times, randomizing different parameters each time, and compared the results to the ratings of the expert.

Step 5: Analyzing the Data via a Machine Learning Algorithm
Machine learning algorithms are developed by giving the machine input values and desired outputs and allowing the computer to find a logical connection between those inputs and outputs. In this way, the computer "learns" to predict the output using other (different) inputs.

To be able to use machine learning, we first had to determine the input parameters. The outputs were the ratings of the experts, collected in step 2.

As input, we decided on three parameters: *Self to Mark, Friends to Mark, Friends Direction*. The first two were calculated in the same way as described for the deterministic algorithm. In this analysis, however, the commander was considered as one of the friendly forces in *Friends to Mark, Friends Direction*, and was not regarded as a separated factor.

The Friends Direction parameter was created by determining the angle between two vectors. The first vector connects the point of the force-mark and the point of the self-mark. The second vector connects the real locations of the force and the subject as had been in the simulated combat (see Figure 5).

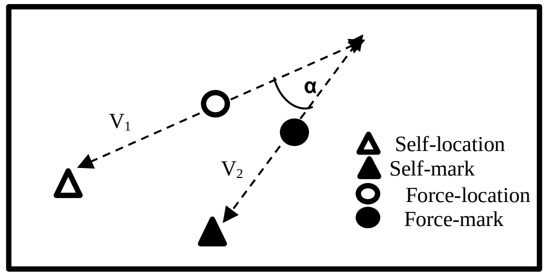

Figure 5. For every friendly force mark, if V_1 is the vector of the actual locations of the force and the subject, and V_2 is the vector between the marks, then the *Friends Direction* parameter is α.

We then used these input factors and the experts' ratings to create two types of machine learning models: a linear regression model and a K-Nearest Neighbors (KNN) classification model. The independent variables formed the input factors and the ratings comprised the expected output.

Linear Regression Model

The creation of the linear regression model included two phases. The first phase involved training the model on 19 of 22 samples; the second phase tested the model on the remaining three samples. Because of the small number of samples, we executed this process 100 times, picking a random 19 samples each time.

During the training process, we created a model of linear regression for each expert and then determined the average rating-error of all the models for each testing sample. We ran this process 100 times and calculated the average of all the individual, rating-error averages.

Although the results were acceptable, we desired them to be more accurate. Accordingly, we decided to run the process again, but with fewer experts. The chosen experts were picked randomly at first, choosing 6-9 of them each time. After that, we tried one last run, in which the relevant experts were chosen by their similarity, using a cluster map chart. As before, we executed this process 100 times for each combination of experts.

KNN Classification Model

K-Nearest Neighbors is a pattern-recognition method, which classifies the input by the k closest training examples. Unlike the regression model, KNN classification cannot predict an exact number. Nevertheless, it is useful for predicting a general rating such as "good" or "bad."

To use the classification model, we grouped the ratings of the experts, as follows:

- 0-25: Very poor
- 26-50: Poor
- 51-75: Good
- 76-100: Very good

Using a KNN algorithm, we created a model that could predict the expert’s rating to a sample in terms of "good" or "poor." This algorithm positions the input data in a multiple-dimensions space and determines to which group a point belongs by examining the closest known points. As we did for the linear regression model, we executed the learning process 100 times, using different sets of samples each time.

RESULTS

Professional Analysis

The results of the first analysis method, the professional manual analysis, were compared in different ways (see Table 1).

Table 1. Ratings of SA by professional analysis, in different types of comparison.

Comparison Type	Groups	Friendly SA rating (average, std)	Enemy SA rating (average, std)
Rank	**Platoons' commanders**	4.9, 2.2	5.5, 2.0
	Companies' commanders	6.1, 1.8	6.0, 1.7
Force type	**Armored**	5.8, 1.9	4.0, 0.0
	Infantry	5.1, 2.2	5.8, 1.9
Situation type	**Routine**	5.0, 2.3	5.3, 2.5
	Before contact	5.5, 1.7	5.0, 0.8
	After contact	5.5, 1.1	6.3, 0.5
	After death	5.3, 2.0	6.4, 1.2
Terrain type	**Open space**	4.8, 2.5	7.0, 0.0
	Urban terrain outskirts	5.4, 1.9	5.0, 1.4
	Urban terrain	5.1, 2.1	5.8, 2.1

From the analysis of these results, some conclusions can be drawn. For example, the armored forces got better ratings than the infantry forces, which can be explained by the experience of the armored forces with maps. Similar explanation can be the reason of the difference between the rating of the platoons' commanders and the companies' commanders.

The Deterministic Algorithm

While running and randomizing the weights and grade scales, every result was tested by comparing the ratings it provided against the ratings which were given by the expert. After all of the runs had finished, we managed to get a final grades scale and a final weight for each factor by selecting those which resulted the closest ratings to the expert's ratings.

When using the best weights and grades-scale, on a scale of 1-10, the average difference between the algorithm's ratings and the SME's ratings was 1.3, with a standard deviation of 0.9.

Table 2 shows the final grades scales. Each grades scale is divided into ranges of error and their grade ranges. The grades for each range is in a linear relationship to the range values.

Table 2. Distance Error and Direction Error Grade Scale

	Error range	Grade range
Distance Error	0-15 meters	70-100
	15-50 meters	60-70
	50-120 meters	30-60
	120-200 meters	10-30
	200< meters	0
Direction Error	0-5 degrees	90-100
	5-30 degrees	50-90
	30-60 degrees	30-50
	60-90 degrees	20-30
	90< degrees	0

The next formula is based on the final weights of the factors:

$$R = 0.4 \times S_l + 0.1 \times F_l + 0.15 \times F_d + 0.1 \times C_l + 0.15 \times C_d + 0.1 \times P$$

In the formula above, R is the SA rating of the subject. S_l is the distance between his location and where he thought himself to be. F_l and F_d are the average distance error and the average direction error between where friendly forces were and the subject marks, respectively. C_l and C_d are the distance and angle between the actual location of the subject's commander and his mark location. P is the percentage of marks the subject used out of all the marks he needed to place. This formula can be used as-is to calculate one's SA, assuming the measured factors are according to those in the formula.

The Machine Learning Algorithm

As mentioned, the final linear regression model was compared to the experts' ratings, both the entire group of experts as well as subsets. For more accurate results, the experts were asked to give their ratings on a scale of 1-100, though scales of 1-7 and 1-10 are usually more popular in SA measuring. Table 3 shows the results on the scales of 1-10 and 1-5.

Table 3. Results of the Linear Regression Model Tests

Test type	Average of errors averages (scale 1-10)	Average of standard deviations (scale 1-10)	Average of errors averages (scale 1-5)	Average of standard deviations (scale 1-5)
All experts	0.82	0.55	0.41	0.275
9 experts	0.88	0.62	0.44	0.31
8 experts	0.81	0.54	0.405	0.27
7 experts	0.64	0.59	0.32	0.295
6 experts	0.67	0.52	0.335	0.26
Filtered by cluster map	0.76	0.56	0.38	0.28

It can be concluded that using all experts' evaluations on a 1-10 scale established a model which can predict the rating of a sample, with average error of less than 1 and standard deviation of 0.55. Meaning, the model can accurately estimate the rating of a sample which would be given by all of the experts, and predicts with increased accuracy the average rating of a sample when considering about half of the experts' ratings.

The KNN classification model was also compared to the results of the experts' rating process (Table 4). The model predicted ratings in the lower grade groups, with accuracy of 57.7% for the "very poor" group and 35.4% for the "poor" group; however, the model assigned a large number of the samples to the lower grade groups, even for those with a better rate from the experts.

Table 4. The percentage of samples in each group that were predicted to be in each of the groups. Correct predictions are those in the bolded cells.

		Samples groups			
		Very poor	Poor	Good	Very good
Prediction	Very poor	57.7%	51.3%	30.3%	28.4%
	Poor	31.6%	35.4%	51.9%	27.6%
	Good	3.6%	8.0%	17.8%	37.4%
	Very good	7.1%	5.2%	0.0%	6.6%

In contrast to the linear regression model, these results did not improve when using a smaller subset of experts.

CONCLUSIONS

The original goal of our work was developing a new method to overcome several challenges in the measurement of situational awareness. The new method needed to provide an objective, quantitative index, which represented an SA rating based on subject's actual knowledge of his surroundings, with the least impact on the simulated combat. This method also required low usage of resources to allow simple application on different scenarios.

The professional analysis was the most detailed method, allowing us to investigate SA by considering a variety of factors. These results, however, were presented only after a few days of analysis by the professional; meaning that, although thorough, this method of analysis requires time and human resources and is not useful for getting immediate answers. In addition, its results follow only one expert's opinion, and can only be considered objective.

It can be concluded that when simplifying SA to its basic factors and focusing on the first level of Endsley's model, the algorithms we developed can fulfil the need of measuring one's mental image. They supply a quick and accurate way to translate comparisons between marks and "ground-truth" into a fixed rating. Thus, they can be merged into different systems to allow automatic calculation of one's answers to SA testing.

The deterministic algorithm provides an explicit calculation of an SA rating. It can also be adjusted to meet one's interpretation of the importance of SA factors. For example, the weight of each factor can be changed if the knowledge of self-location is considered less important than understanding how far away other troops are located.

Each of the machine-learning models showed different results. While the linear regression model predicted one's SA rate accurately (when relating more than half of the experts), the KNN model was not able to predict the results of samples rated highly by the experts. Apparently, the KNN model requires more varied ratings for training process, unlike our data, which included mostly low-rated samples.

It can be concluded that QIRT allows the creation of pre-made models, which can be used for assessing one's SA while in action. Since its ratings depend only on locations and marks, the data collection process can be made with little interference. Its creation requires one-time usage of resources, and once the model is created, it can be applied in varied scenarios. Such model's result is a simple, comparable rate, which can be used for further analysis and comparison between different participants.

Future Plans

The main ability the QIRT method provides is the in-action calculation of an SA rating. When testing a new C4I system or in a trainer, using either the deterministic algorithm or the machine learning models in simulations can be helpful for measuring the influence on a subject immediately. Moreover, applying these algorithms makes the SA rating a simple, understandable number, so anyone can understand its validity. Such rating methods are best suited, therefore, for testing SA in systems with assorted activities.

One primary ability lacking in the presented models is the ability to measure the commander's knowledge of the enemy situation. Unlike friendly forces, the enemy situation image is based on different factors, and probably needs to be analyzed differently, as well. Once the relevant factors are determined, QIRT can be used to develop models for measuring enemy SA, too.

Another issue is the situation type factor (i.e., routine action, before combat, etc.). In the professional analysis, the SME considered the different situation types when rating the samples, which are critical to the commander's ability and necessity to pay attention to different SA elements. For example, on routine action the commander can easily follow the positions of his troops and their fitness for combat; however, contact with the enemy leads to sudden and unpredictable changes of the forces status. It may be more important, however, to know where friendly forces are after a combat encounter than while on routine movement. The situation type might influence an experts' opinion on an SA rating, and therefore be more relevant for the models to get more accurate prediction.

REFERENCES

Endsley, M.R. & Garland, D.J. (Eds.) (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M.R (1995). Measurement of situation awareness in dynamic systems. *Human Factors 37 (1): 65–84.*

Endsley, M.R. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 82–86). Santa Monica, CA: The Human Factors and Ergonomics Society.*

Flin, R. & O'Connor, P. (2001). Applying crew resource management in offshore oil platforms. *Improving teamwork in organization: Applications of resource management training (pp. 217–233). Hillsdale, NJ: Erlbaum*

Gaba, D.M.; Howard, S.K.; Small, S.D. (1995). "Situation awareness in nesthesiology". *Human Factors. 37 (1): 20–31*

Salmon, P. & Stanton, N. & Walker, G. & Green, D. (2006). Situation awareness measurement: a review of applicability for C4i environments. *Journal of Applied Ergonomics 37*

Shamash E. M. (2017). The Effect of Physical Dispersion on the Quality of the Operational Picture in Crisis Management Teams. *University of Haifa*, Haifa, Israel.

Taylor, R.M., (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations*, Neuilly Sur Seine, France.