7-2012

Adaptive CGF for pilots training in air combat simulation

Teck-Hou TENG

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Wee-Sze ONG

Kien-Lip LEE

Citation

TENG, Teck-Hou; TAN, Ah-hwee; ONG, Wee-Sze; and LEE, Kien-Lip. Adaptive CGF for pilots training in air combat simulation. (2012). *Proceedings of the 15th International Conference on Information Fusion, Singapore, Jul 9-12*. 2263-2270.

Available at: https://ink.library.smu.edu.sg/sis_research/6663

Adaptive CGF for pilots training in air combat simulation

Teng, Teck-Hou; Tan, Ah-Hwee; Ong, Wee-Sze; Lee, Kien-Lip

2012

Teng, T.-H., Tan, A.-H., Ong, W-S., & Lee, K.-L. (2012). Adaptive CGF for pilots training in air combat simulation. 2012 15th International Conference on Information Fusion (FUSION), 2263 - 2270.

https://hdl.handle.net/10356/101744

Adaptive CGF for Pilots Training in Air Combat Simulation

Teck-Hou Teng*, Ah-Hwee Tan*, Wee-Sze Ong† and Kien-Lip Lee‡
*School of Computer Engineering, Nanyang Technological University
†DSO National Laboratories
‡CAE Singapore (S.E.A.) Pte. Ltd.

***Abstract*—Training of combat fighter pilots is often conducted using either human opponents or non-adaptive computer-generated force (CGF) inserted with the doctrine for conducting air combat mission. The novelty and challenges of such non-adaptive doctrine-driven CGF is often lost quickly. Incorporating more complex knowledge manually is known to be tedious and time-consuming. Therefore, a study of using adaptive CGF to learn from the real-time interactions with human pilots to extend the existing doctrine is conducted in this work. The goal of this study is to show how an adaptive CGF can be more effective than a non-adaptive doctrine-driven CGF for simulator-based training of combat pilots. Driven by a family of self-organizing neural network, the adaptive CGF can be inserted with the same doctrine as the non-adaptive CGF. Using a commercial-grade training simulation platform, two human-in-the-loop (HIL) experiments are conducted using the adaptive CGF and the non-adaptive doctrine-driven CGF to engage two diverse groups of human pilots in 1-v-1 dogfights. The quantitative results and qualitative assessments of the CGFs by the human pilots are collected for all the training sessions. The qualitative assessments show the trainee pilots are able to match the adaptive CGF to the desirable attributes while the veteran pilots are only able to observe some learning from the adaptive CGF. The quantitative results show that the adaptive agent needs a lot more training sessions to learn the necessary knowledge to match up to the human pilots.**

I. Introduction

There are numerous challenges to ensure the efficient and accurate transfer of human-level knowledge into autonomous knowledge-based system [1]. Often, the knowledge engineers are not the subject matter experts and vice versa. This inability to correctly transfer the expert knowledge into the expert systems [2] is leading to the irrecoverable lost of rich sources of knowledge, severely limiting the applications of existing expert systems [3].

Many works in artificial intelligence [4] use human as their main source of inspiration. After the Deep Blue supercomputer [5], the IBM Watson was created to, again, successfully challenge the human experts in their area [6]. The approach used in these works is to learn from human the knowledge that can be used against them. Such leveling up of knowledge serves to further the intellectual horizon of the human counterparts. One such area of application is in the simulator-based training of human pilots for combat missions [7]. The non-deterministic adaptation of the adaptive CGF to the evolving situations has much greater training value to the trainee pilots than a non-adaptive doctrine-driven CGF.

This work builds on an earlier work [8] to learn counter strategies in 1-v-1 air combat scenario. The learning of counter strategies was earlier carried out against a non-adaptive doctrine-driven CGF. Inserted with the same doctrine, learning was shown converging to the winning air combat maneuvers after a brief encounter. In this work, learning of the counter air combat maneuvers is similarly carried out using a fusion architecture for learning and cognition (FALCON) integrated with a temporal difference (TD) method known as the TD-FALCON [9]. Using reinforcement learning, it adapts during its real-time interactions with the environment. In this work, the convergence threshold is significantly raised by inviting the participation of the human pilots instead of using the non-adaptive rule-based CGF for the same learning task.

Two HIL experiments were conducted using a group of trainee pilots and a group of veteran combat pilots. The pilots engage in 1-v-1 air combat against either an adaptive CGF or a non-adaptive doctrine-driven CGF. The identity of the CGFs are hidden from the pilots as it is part of this study to find out whether the pilots are able to match the adaptive CGF with the desirable attributes. From the 1^{st} HIL experiment, the adaptive CGF showed some amount of adaptation to score a temporary advantage over the trainee pilots. The adaptive CGF was also correctly matched to the desirable attributes by the trainee pilots in their qualitative assessments of the CGFs. Using slightly changed conditions for the 2^{nd} HIL experiment, the veteran combat pilots were able to observe some amount of adaptation by the adaptive CGF. But the ongoing adaptation of the adaptive CGF to the unfamiliar conditions led the veteran pilots to feel it is not learning fast enough.

The rest of the paper is organized as follows. A survey of the related works is presented in Section II. A summary of the learning strategy of TD-FALCON is presented in Section III. This is followed by the descriptions of the 1-v-1 air combat scenario between the CGFs and the pilots in Section IV. Details on how the HIL experiment and the questionnaires are designed are provided in Section V. This is followed by the presentation of the execution details of the HIL experiments and also the analysis and the discussions of the experimental results in Section VI. Last but not least, the implications arising from this work and the future works to build on this work are presented in Section VII.

II. Related Work

This work is an exploratory study on using adaptive CGF to learn counter air combat strategies to the combat pilots while at the same time providing some training value to the air combat pilots. Some amount of knowledge-based techniques have long been used in various aspects of the simulators [10]. Artificial intelligence techniques were specifically used in the earlier training simulators for improving the user experience [11]. Improvement in the air combat ability of the trainee pilots against computer-driven adversary was also reported [7].

An earlier use of neural network technique to adapt the helicopter system to the learning curve of the trainee pilot during training was outlined [12]. Another recent work used a bio-inspired immune system paradigm with a neural flight controller to select and construct air combat maneuvers for 1-v-1 air combat maneuvering [13]. However, there is no direct participation of the human pilots similar to this work. In another work, genetic algorithm demonstrated using small conventional fighting units to build large formation tactics. Several offline simulations were conducted using large aircraft formations. Though the developed tactics are compatible to the existing combat principles, the effectiveness of the discovered air combat tactics against human pilots in real-time flying remains unvalidated.

Several knowledge acquisition and elicitation strategies for building up expert systems were proposed [1]. The STRADS program is an example of discovering knowledge for potential combat scenarios through simulation [14]. Learning by observation [15] and by interactions [16] remain some of the popular approaches in learning directly from a human. Similar to the concept of *learning in the wild* [17], this work does not require any extra effort from the human pilots for learning counter air combat strategies. The pilots are only required to focus on flying their virtual aircraft to the best of their ability. However, learning in [17] is conducted on e-mail communications between humans, not on the more complex learning of counter strategies.

The dynamic and fluid nature of air combat is not suitable for the use of any artificial maieutic strategy such as the one reported in [18]. The learning approach used in this work is for learning during the real time interactions between two adversarial entities. In this respect, the Deep Blue supercomputer and the DeepQA project are similar. However, the Deep Blue supercomputer only succeeded in challenging a chess master using efficient search technique and unrivalled hardware. This leaves the DeepQA project [6] the only work that come close to the online learning approach used in this work. However, it remains unclear how well it scales for the learning of counter air combat maneuvering strategies.

III. THE REINFORCEMENT LEARNING MODEL

The adaptive CGF is driven by a self-organizing neural network known as FALCON. Based on the adaptive resonance theory (ART), it can learn and generalize on the situations incrementally. Using reinforcement learning, knowledge is discovered during real-time interactions with the environment.

A. FALCON Model and Processes

The FALCON network [19] employs a 3-channel architecture (Fig. 1), comprising of a category field F_2^c and three input fields, namely a sensory field F_1^{c1} for representing current states, an action field F_1^{c2} for representing actions, and a reward field F_1^{c3} for representing reinforcement values. A brief summary of the FALCON generic network dynamics, based on fuzzy ART operations [20], is described below.

Input vectors: Let $\mathbf{S} = (s_1, s_2, \ldots, s_n)$ denote the state vector, where $s_i \in [0,1]$ indicates the sensory input i. Let $\mathbf{A} = (a_1, a_2, \ldots, a_m)$ denote the action vector, where $a_i \in [0,1]$ indicates a possible action i. Let $\mathbf{R} = (r, \bar{r})$ denote the reward vector, where $r \in [0,1]$ is the reward signal value and $\bar{r}$ (the complement of r) is given by $\bar{r} = 1 - r$.

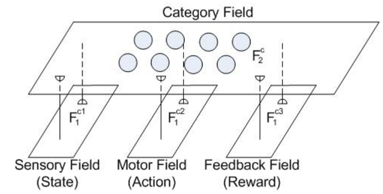


Fig. 1. The FALCON Architecture.

Complement coding is used to normalize the magnitude of the input vectors to prevent the code proliferation problem.

Activity vectors: Let $\mathbf{x}^{ck}$ denote the F_1^{ck} activity vector for $k = 1, \ldots, 3$. Let $\mathbf{y}^c$ denote the F_2^c activity vector. Upon input presentation, $\mathbf{x}^{c1} = \mathbf{S}$, $\mathbf{x}^{c2} = \mathbf{A}$, and $\mathbf{x}^{c3} = \mathbf{R}$.

Weight vectors: Let $\mathbf{w}_j^{ck}$ denote the weight vector associated with the jth node in F_2^c for learning the input patterns in F_1^{ck} for $k = 1, \ldots, 3$.

Parameters: The FALCON's dynamics is determined by choice parameters $\alpha^{ck} > 0$ for $k = 1, \ldots, 3$; learning rate parameters $\beta^{ck} \in [0,1]$ for $k = 1, \ldots, 3$; contribution parameters $\gamma^{ck} \in [0,1]$ for $k = 1, \ldots, 3$ where $\sum_{k=1}^{3} \gamma^{ck} = 1$; and vigilance parameters $\rho^{ck} \in [0,1]$ for $k = 1, \ldots, 3$.

Code activation: A bottom-up propagation process first takes place in which the activities (known as choice function values) of the cognitive nodes in the F_2^c field are computed. Specifically, given the activity vectors $\mathbf{x}^{c1}$, $\mathbf{x}^{c2}$ and $\mathbf{x}^{c3}$ (in the input fields F_1^{c1}, F_1^{c2} and F_1^{c3} respectively), for each F_2^c node j, the choice function T_j^c is computed as follows:

$$T_j^c = \sum_{k=1}^{3} \gamma^{ck} \frac{|\mathbf{x}^{ck} \wedge \mathbf{w}_j^{ck}|}{\alpha^{ck} + |\mathbf{w}_j^{ck}|}$$

where the fuzzy AND operation $\wedge$ is defined by $(\mathbf{p} \wedge \mathbf{q})_i \equiv min(p_i, q_i)$, and the norm $|.|$ is defined by $|\mathbf{p}| \equiv \sum_i p_i$ for vectors $\mathbf{p}$ and $\mathbf{q}$.

Code competition: A code competition process follows under which the F_2^c node with the highest choice function value is identified. The winner is indexed at J where

$$T_J^c = \max\{T_j^c : \text{for all } F_2^c \text{ node } j\}$$

When a category choice is made at node J, $y_J^c = 1$; and $y_j^c = 0$ for all $j \neq J$. This indicates a winner-take-all strategy.

Template matching: Before node J can be used for learning, a template matching process checks that the weight templates of node J are sufficiently close to their respective activity patterns. Specifically, resonance occurs if for each channel k, the *match function* m_J^{ck} of the chosen node J meets its vigilance criterion:

$$m_J^{ck} = \frac{|\mathbf{x}^{ck} \wedge \mathbf{w}_J^{ck}|}{|\mathbf{x}^{ck}|} \geq \rho^{ck}$$

If any of the vigilance criteria is violated, mismatch reset occurs. The choice function T_J^c is set to 0 for the duration of the input presentation. Another F_2^c node J using a revised vigilance criterion until a resonance is achieved. This search and test process is guaranteed to end as FALCON will either find a *committed* node that satisfies the vigilance criterion or activate an *uncommitted* node which would definitely satisfy the vigilance criterion due to its initial weight values of $1s$.

Template learning: Once a node J is selected, for each channel k, the weight vector $\mathbf{w}_J^{ck}$ is modified by the following learning rule:

$$\mathbf{w}_J^{ck(\text{new})} = (1 - \beta^{ck})\mathbf{w}_J^{ck(\text{old})} + \beta^{ck}(\mathbf{x}^{ck} \wedge \mathbf{w}_J^{ck(\text{old})})$$

For an uncommitted node J, the learning rates β^{ck} are typically set to 1. For committed nodes, β^{ck} can remain as 1 for fast learning or below 1 for slow learning in a noisy environment. When an uncommitted node is selecting for learning, it becomes *committed* and a new uncommitted node is added to the F_2^c category field.

B. Incorporating Temporal Difference Method

Outlined in Fig. 2, the TD-FALCON algorithm [9] incorporates Temporal Difference (TD) methods to estimate and learn the value functions of state-action pairs $Q(s,a)$ that indicates the goodness for taking a certain action a in a given state s. This is learned as the feedback signal and is used in the selection of the action choices.

1: Initialize FALCON
2: Sense the environment and formulate a state representation s
3: Use *Action Selection Policy* to decide between **Exploration** and **Exploitation**
4: **if** Exploration **then**
5: Use *Exploration Strategy* to select an action choice from action space
6: **else if** Exploitation **then**
7: Use *Direct Code Access* to select an action choice from existing knowledge [21]
8: **end if**
9: Use action choice a on state s for state s'
10: Evaluate effect of action choice a to derive a reward r from the environment
11: Estimate the Q-value function $Q(s,a)$ following a temporal difference formula given by $\Delta Q(s,a) = \alpha TD_{err}$
12: Present state S, action A and reward R vectors for **Adaptation**
13: Update the current state $s = s'$
14: Repeat from Step 2 until s is a terminal state

Fig. 2. The TD-FALCON Algorithm

Iterative Value Estimation: A value function based on a temporal difference method known as Bounded Q-Learning is used to iteratively estimate the value of applying action choice a to situation s. The estimated Q-value $Q(s,a)$ is learned by TD-FALCON during reinforcement learning. The temporal difference of the value function is iteratively estimated using

$$\Delta Q(s,a) = \alpha TD_{err}(1 - Q(s,a))$$

where $\alpha \in [0,1]$ is the learning parameter, the term $(1 - Q_j(s,a))$ allows the adjustment of Q-values to be self-scaling in such a way that it will not be increased beyond 1.0 and TD_{err} is the temporal error term which is derived using

$$TD_{err} = r + \gamma \max_{a'} Q(s',a') - Q(s,a)$$

where $\gamma \in [0,1]$ is the discount parameter and the $\max_{a'} Q(s',a')$ is the maximum estimated value of the next state s' and r is either the intermediate or terminal reward.

C. Adaptive ϵ-Greedy Action Selection Policy

Using this policy, exploration is performed with a probability of ϵ where $\epsilon \in [0,1]$ [22]. An interval success rate ϕ which is derived using $\phi = \frac{w_s}{w_n}$ where w_s is the number of successful trials within w_n training iterations is used to revise ϵ as $1-\phi$ after every w_n training iterations. The revised ϵ is linearly decayed over the next w_n training iterations using an ϵ-decay rate δ which is derived using $\frac{\epsilon}{w_n}$. Such an approach gradually increases exploitation of the learned knowledge within w_n training iterations. This allows it to incrementally evaluate the effectiveness of the learned knowledge for the situations.

IV. 1-v-1 Air Combat between the CGFs and the Human Pilots

The problem domain in this work is based on a classical 1-v-1 pursuit-evasion problem in three-dimensional airspace [23]. The CGF controls the virtual entity named FalconX while the pilot controls the virtual entity named CGFPlayer. Both of them are tasked to out-maneuver each other to enter into a position to eliminate each other using AIM-9 missiles. Unlike [8], this work investigates how well an adaptive CGF can respond to the human pilots as compared to a non-adaptive CGF in 1-v-1 dogfights.

A. The Computer-Generated Forces

An adaptive CGF and a non-adaptive doctrine-driven CGF, driven separately by the same ART-based self-organizing neural network known as TD-FALCON, are used in this work. The state space and the action space for the CGFs and the evaluative feedback used by the adaptive CGF are described.
The State Space: The state space is automatically extracted from the air combat maneuver (ACM) doctrine using the technique outlined in [8]. As illustrated in Fig. 3, the extracted state space is a fusion of *relative parameters* such as range and angular position, own *entity ACM parameters* such as current maneuver, maneuver lock status and own *entity parameters* such as altitude and air speed. A total of 15 attributes are extracted from the ACM doctrine.

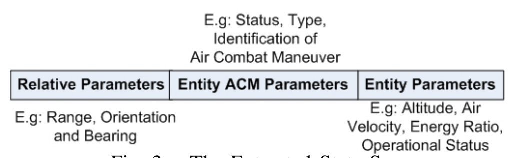


Fig. 3. The Extracted State Space

The Action Space: The CGFs have to identify the effective air combat maneuvers to out-maneuver the pilots to get into a good position to fire the AIM-9 missile at the opponent. A combination of 13 defensive and offensive air combat maneuvers are available to the CGFs during the dogfights. Unlike the pilots, the CGFs cannot control the execution of the pre-defined air combat maneuvers.
Evaluative Feedback: This is used to steer the reinforcement learning to achieve the desired effect of the learning task. There are the intermediate reward for the intermediate states and the terminal reward for the terminal states. The intermediate reward communicates the effect of the chosen response at the intermediate states. The terminal reward quantifies the outcome at the terminal states. Details on these two types of evaluative feedback can be found in [8].

B. The Human Pilots

Two groups of pilots from different air force organisations are invited for the HIL experiments. The first group of pilots comprises of two trainee pilots and the second group of pilots comprises of three veteran pilots. Such an arrangement allows a broader assessment of the CGFs.
Trainee Pilots: The first group of participants are trainee RSAF (Republic of Singapore Air Force) pilots in their 20s. They had completed their basic military training stint and were specialising to fly the fighter jets. They had no *prior* experience on the use of the commercial-grade simulation platform used in this work. However, almost all of them have some experience with flying in some game-like flight

simulators. Studies have shown that this can translate to better performance when it comes to actual flight [24].

Veteran Pilots: The second group of participants are veteran CAF (Canadian Air Force) combat pilots. They had served their country in actual air combat missions as fighter jet pilots. They have retired from regular services and are the subject matter experts (SMEs) to a multinational Canadian-based simulator manufacturer CAE®Inc. Being the SMEs to the simulator products used in the HIL experiment, the veteran pilots are familiar with the handling of the modeled aircraft.

V. The Human-In-the-Loop (HIL) Experiment

A design methodology of the HIL experiment to ensure sufficient amount of quantitative and qualitative data is collected is presented over here. Also included is the design of the trial and session questionnaires used for collecting qualitative assessments of the CGFs by the human pilots.

A. Designing the HIL Experiments

Unlike an earlier work [8] where the experiments are conducted between two CGFs, human pilots are invited to fly against the CGFs in the HIL experiments. The human factor [25] raises the need for a well-considered set of timing parameters. Therefore, the following design methodology is used to derive these timing parameters.

Let **C** denotes a set of CGFs and **P** denotes a set of pilots. A continuous 1-v-1 air combat between a CGF where $CGF \in \mathbf{C}$ and a pilot P where $P \in \mathbf{P}$ is taken to be a sortie. The sortie terminates when either the CGF or the pilot P is eliminated or when neither of them can eliminate each other after a predetermined duration T_{sortie} minutes (advised by the SME).

Let $\mathbf{C}_{nf}(P)$ be the set of CGFs who has not flown against pilot P and $\mathbf{C}_f(P)$ be the set of CGFs who has flown against pilot P, each trial is comprised of $N_{sorties}$ sorties between a pilot P and a CGF where $CGF \in \mathbf{C}_{nf}(P)$. Pilot P is required to fly $N_{sorties}$ number of sorties with each CGF and it can be derived using

$$N_{sorties} = \left\lfloor \frac{T_{alert}}{T_{sortie}} \right\rfloor \quad (1)$$

where T_{alert} is an estimated duration (in minutes) the pilot is able to continuously fly the simulated aircraft without any sign of simulator sickness [25]. An intermission of T_{TR} minutes is included for completing a trial questionnaire and to get some amount of rest before the next trial. Therefore, using (1), the expected maximum duration T_{trial} of each trial is derived using

$$T_{trial} = N_{sorties} \times T_{sortie} + T_{TR} \quad (2)$$

After each trial, the CGF is moved from $\mathbf{C}_{nf}(P)$ to $\mathbf{C}_f(P)$. The subsequent trial is for pilot P to engage in 1-v-1 air combat against CGF' where $CGF' \in \mathbf{C}_{nf}(P)$. A session is completed when $\mathbf{C}_{nf}(P) \equiv \emptyset$ and $\mathbf{C}_f(P) \equiv \mathbf{C}$. The intermission at the end of each session is extended by T_{SR} minutes. This is for the completion of a trial questionnaire, a session questionnaire and to get some amount of rest before the next session. Therefore, using (2), the expected maximum duration $T_{session}$ of each session S_x is derived using

$$T_{session} = |\mathbf{C}| \times T_{trial} + T_{SR} \quad (3)$$

Let $\mathbf{P}_{nf}$ be the set of pilots who have not flown and $\mathbf{P}_f$ be the set of pilots who has flown, on the completion of each session, pilot P is moved from $\mathbf{P}_{nf}$ to $\mathbf{P}_f$. A cycle of the HIL experiment is completed when $\mathbf{P}_{nf} \equiv \emptyset$ and $\mathbf{P}_f \equiv \mathbf{P}$. Using (3), the number of cycles N_{cycle} of the HIL experiments can be completed by each pilot P in a day is derived using

$$N_{cycle} = \left\lfloor \frac{T_{day}}{T_{session} \times |\mathbf{P}|} \right\rfloor \quad (4)$$

where T_{day} is the agreed amount of flying time (in minutes) in a day by the pilots. Using this approach ensures all pilots will fly the same number of cycles with all CGF types. Using (4), the total number of sorties $N_{dogfight}(CGF, P)$ a CGF has with each pilot P in a day is derived using

$$N_{dogfight}(CGF, P) = N_{cycle} \times N_{sorties} \quad (5)$$

From (5), the total amount of time pilot P spent flying against a CGF is derived using

$$T_{dogfight}(CGF, P) = N_{dogfight}(CGF, P) \times T_{sortie} \quad (6)$$

The HIL experiment combines the sorties flown by all pilots against the same CGF. In this way, the desired number of sorties $N'_{sorties}(CGF)$ for the CGF can be obtained in half the time. Using (6), the number of days required for the HIL experiment T_{HIL} is derived using

$$T_{HIL} = |\mathbf{C}| \times \frac{N'_{sorties}(CGF) \times T_{sortie}}{T_{dogfight}(CGF, P) \times |\mathbf{P}|}$$

A schedule for the HIL experiment is drawn up using the derived timing parameters. From experience, the HIL experiment is able to complete ahead of such a planned schedule. Smooth execution of the HIL experiment enhances the credibility of the experimental results.

B. The Questionnaires

The quantitative results from the HIL experiment reveals only the final outcome of the sortie. For a more balanced evaluation of the CGF, qualitative assessments are used to capture aspects of the CGFs missed by the quantitative results. The questionnaires are issued to the pilots after each trial and session. The pilots are requested to assess the CGF on the following attribute - *predictable*, *intelligent*, *skillful*, *challenging*, *adaptive* and *aggressive*. The pilots are briefed on the agreed definition of the attributes and the different frequencies of occurrence to minimize any ambiguity in their interpretations. Therefore, the pilots are assumed to conduct their qualitative assessments of the CGFs using a similar level of understanding of the terms.

Trial Questionnaire: After each trial, the pilots are required to complete a *trial questionnaire* comprising of the first five attributes. Given that the requested attributes are qualitative in nature, it is only reasonable and possible for the pilots to conduct their assessments using the following perceived frequencies of occurrence: *never*, *rarely*, *sometimes*, *frequently*, *most of the time* and *always*.

Session Questionnaire: After each session, the pilots are also required to complete a session questionnaire. Unlike the trial questionnaire, the pilot ranks all the CGF types he had flown against during that session using all six attributes. A CGF is ranked more highly when it is showing more of the specific attributes in comparison to the other CGF. During ranking, to conceal their underlying characteristics, the CGFs are presented to the pilots as CGF-A and CGF-B in their order of appearance.

VI. Experiments and Results

Two independent HIL experiments were conducted using the trainee pilots and the veteran pilots. The purpose of the experiments is to study, in comparison to a non-adaptive doctrine-driven CGF, how well the adaptive CGF is adapting against the pilots in 1-v-1 air combat scenario and to also gather qualitative assessments from the pilots on these two types of CGF.

A self-organizing neural network known as TD-FALCON is used to drive the adaptive CGF (L-CGF) and the non-adaptive CGF (S-CGF). The learning mechanism of the TD-FALCON is disabled for the S-CGF. The same air combat doctrine is inserted into the L-CGF and the S-CGF using the technique described in [8]. The parameters identified for the TD-FALCON are presented in Table I. The HIL experiment is designed using $T_{alert} = 75$ minutes, $T_{sortie} = 5$ minutes, $T_{TR} = T_{SR} = 5$ minutes and $T_{day} \geq 480$ minutes.

TABLE I
Parameters of TD-FALCON and Action Selection Policy

DA−FACLON Parameters	
Choice Parameters $(\alpha^{c1}, \alpha^{c2}, \alpha^{c3})$	0.1,0.1,0.1
Learning Rates $(\beta^{c1}, \beta^{c2}, \beta^{c3})$	1.0,1.0,1.0
Contribution Parameters $(\gamma^{c1}, \gamma^{c2}, \gamma^{c3})$	0.33,0.33,0.33
Perform Vigilance $(\rho_p^{c1}, \rho_p^{c2}, \rho_p^{c3})$	0.0,0.0,0.45
Learn Vigilance $(\rho_l^{c1}, \rho_l^{c2}, \rho_l^{c3})$	0.0,1.0,0.45
Temporal Difference Learning Parameters	
Learning Rate α	0.5
Discount Factor γ	0.1
Initial Q-Value	0.5
ϵ-Greedy Policy Parameters	
Initial ϵ Value	0.0
ϵ Decay Rate	0.0005
Window Size w_n	5

A. Descriptions of the HIL Experiment

The physical set-up and the execution procedures of the HIL experiment are described here. The specific handling of the CGFs during the HIL experiment is also included as part of the execution procedures.

Physical Set-Up: The HIL experiment is conducted using the flight training simulator set-up illustrated in Fig. 4. The physical set-up comprises one unit of desktop workstation WS1, two units of desktop computer DC1 and DC2 and a pair of joysticks to simulate the HOTAS in the cockpit of a fighter jet. The joysticks are connected to the desktop computer DC1 running the commercial-grade simulator software known as STRIVE CGF®Studio. The dynamics and physics of flying the simulated aircraft within the three-dimensional airspace and the effects of the missiles launched from the simulated aircraft are modeled using the STRIVE CGF®Studio.

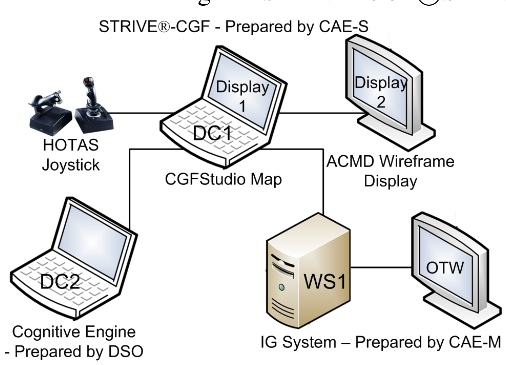


Fig. 4. An illustration of the physical set-up of the HIL experiment.

From Fig. 4, DC1 is connected to the desktop computer DC2 running the cognitive engine in a client-server configuration illustrated in Fig. 5. The ICON®interface facilitates the communication between DC1 and DC2. DC1 is also connected to the desktop workstation WS1 designated as the Image Generator (IG) for generating the Out-of-The-Window (OTW) view used by the pilots. It can generate a multi-channel three-dimensional panoramic OTW view of 100^o horizontal and 56^o vertical of the airspace. A head-up display (HUD) unit is superimposed onto the OTW view to provide an artificial horizon, the altitude, the rate of climb, the air speed, the G-meter and the compass heading of the aircraft. The weapons aiming system is omitted from the HUD.

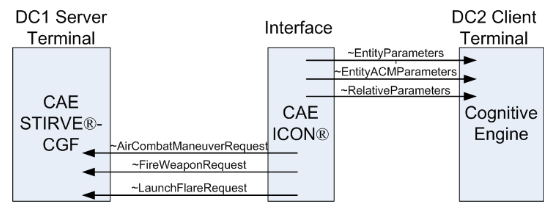


Fig. 5. An illustration of the client-server set-up between the AI terminal and Simulator terminal

Execution Procedures: A team comprising of the pilot, the *co-pilot* and the support crews are necessary for the HIL experiment. The pilot does the actual flying of the simulated aircraft to engage in 1-v-1 dogfights with a CGF. Due to the restricted view of the airspace arising from using a single channel IG, the *co-pilot* provides additional information such as the elevation and general position of the opponent when it is not in the OTW view. The *co-pilot* gathers such information using the ACMD wireframe display and the CGFStudio Map in DC1 (see Fig. 4).

In addition, the *co-pilot* also records the number of missiles fired and the final outcome of the sortie using a recording form. Any other observations are included as remarks in it as well. The recorded number of missile fired indicates the level of difficulty in engaging the CGFs during the 1-v-1 dogfight. However, it is omitted from the experimental results presented in this paper. The recordings of the final outcome is used as the quantitative results of the HIL experiment. The other observations are recorded for a better understanding of the actual proceedings of the sorties.

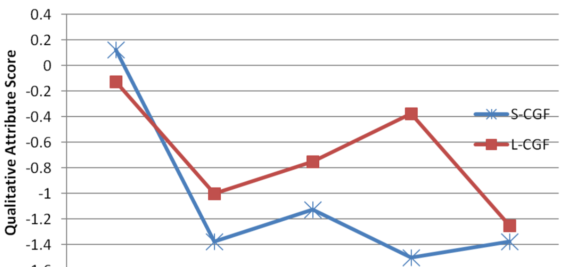
The support crews comprising of the AI team and the simulator team ensure the smooth execution of the HIL experiment. The simulator team is needed to make speedy recovery of the HIL experiment when there is any technical glitches. The pilot flies against either the L-CGF or the S-CGF during a trial. On top of not revealing the identity of the CGF to the pilots, the order of appearance of the CGFs during the session is also randomized. To collect the same amount of experimental data from all the pilots with all the CGFs, the AI team needs to ensure the correct CGF is used for the trials.

Both CGFs are inserted with the same doctrine at the very first sortie. Only the L-CGF updates its knowledge base by learning its interactions with the human pilots. The AI team is required to ensure the updated knowledge base is used for the subsequent trials involving the L-CGF with any of the pilots. The AI team is also required to ensure the correct set of initial conditions is used for the trials when there is more than one set of initial conditions.

B. Air Combat with the Trainee Pilots

Two trainee pilots took turns to be the pilot and the *co-pilot* in this HIL experiment. Like [8], a *gun-fight* scenario modelled using the initial conditions illustrated in Fig. 6 is used. This HIL experiment is comprised of six sessions of two trials per session and 15 sorties per trial. Each trial contributes a quantitative data point and qualitative data point.

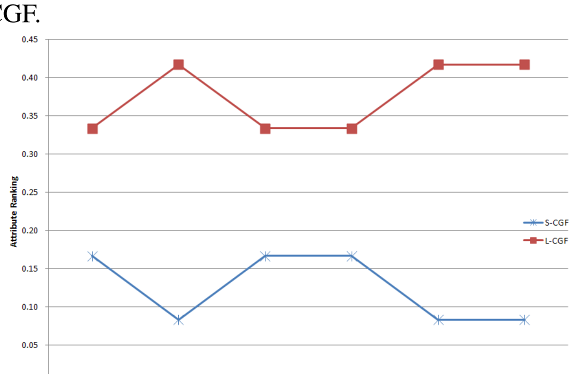
Fig. 6. An illustration of the *gun-fight* initial conditions used for the HIL experiments with the trainee pilots



	Predictable	Intelligent	Skilful	Challenging	Adaptive
S-CGF	0.125	-1.375	-1.125	-1.5	-1.375
L-CGF	-0.125	-1	-0.75	-0.375	-1.25

Fig. 7. Comparisons of the qualitative assessments of the L-CGF and the S-CGF by the trainee pilots using the trial questionnaire (higher score means higher perceived rate of occurrence).

Qualitative Data: From the plot of aggregated numerical score of the qualitative assessments in Fig. 7, both trainee pilots unanimously rated the S-CGF to be more predictable than the L-CGF. As gathered from the same set of trial questionnaires, the L-CGF is also perceived to be more intelligent, more skillful, more challenging and more adaptive than the S-CGF.



	Predictable	Intelligent	Skilful	Challenging	Adaptive	Aggressive
S-CGF	0.17	0.08	0.17	0.17	0.08	0.08
L-CGF	0.33	0.42	0.33	0.33	0.42	0.42

Fig. 8. Direct comparisons of the L-CGF and the S-CGF by the trainee pilots using the session questionnaire (higher score means higher perceived rate of occurrence).

From Fig. 8, direct comparison between the L-CGF and the S-CGF reveal mostly similar viewpoints both trainee pilots have of the CGFs. However, both trainee pilots perceived the L-CGF to be more predictable than the S-CGF when comparing them directly. Another observation is the gap between the CGFs is wider in a direct comparison than when the CGFs are assessed individually.

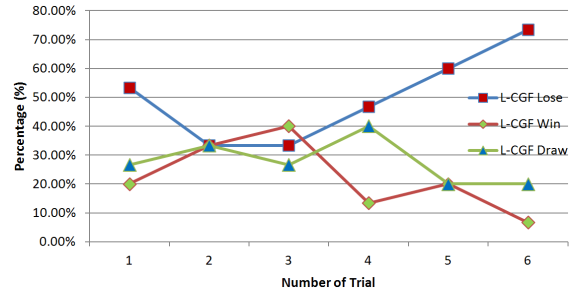


Fig. 9. Plots of the quantitative data from the 1-v-1 dogfights between the L-CGF and the trainee pilots

Quantitative Data: From Fig. 9, the L-CGF is observed drawing more and winning less at its 1^{st} trial with the trainee pilots. This is followed by an upward trend of L-CGF winning both trainee pilots more at the 3^{rd} trial. Using the learned knowledge, the L-CGF is still able to draw more with both trainee pilots at the 4^{th} trial. However, both trainee pilots caught up with the L-CGF from the 5^{th} trial onwards.

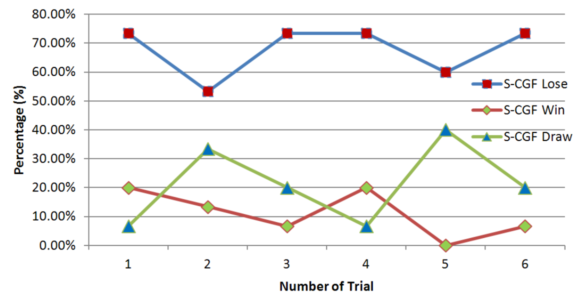


Fig. 10. Plots of the quantitative data from the 1-v-1 dogfights between the S-CGF and the trainee pilots

In contrast, as shown in Fig. 10, the non-adaptive S-CGF is losing to both trainee pilots for more than 50% in most of the trials. At other times, it appears to be drawing more with the trainee pilots. Given that the S-CGF is non-adaptive and the same initial conditions are used for all the trials, the fluctuations of the plots in Fig. 10 is due to the ongoing adaptation of the trainee pilots. This fact is slightly masked in Fig. 9 as the L-CGF is also adapting to the trainee pilots in those trials.

C. Air Combat with the Veteran Pilots

This HIL experiment involved three veteran pilots engaging in 1-v-1 dogfights with the L-CGF and the S-CGF. The same set of experiment parameters presented in Table I is used here. However, the initial conditions are fundamentally changed to reflect a *missile-fight* scenario rather than a *gun-fight* scenario. This change is recommended by the veteran pilots because the AIM-9 missiles rather than the aircraft gun are used to eliminate the opponents. Therefore, two new sets of initial conditions illustrated in Fig. 11 are used. Also, unlike the preceding HIL experiment, this HIL experiment is comprised of nine sessions. Due to the changing conditions during this HIL experiment, the number of trials flew by each veteran pilots are different.

Qualitative Data: From the plot of the aggregated numerical scores of the qualitative assessments in Fig. 12, three veteran pilots perceived the L-CGF to be more predictable, less intelligent, less skillful, less challenging but more adaptive than the S-CGF. The gap between the intelligent, skillful, challenging and adaptive attributes are smaller compared to

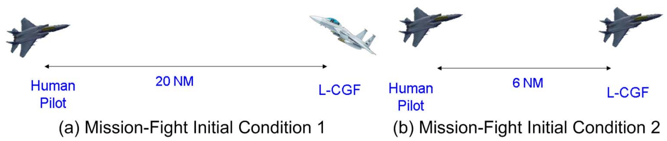
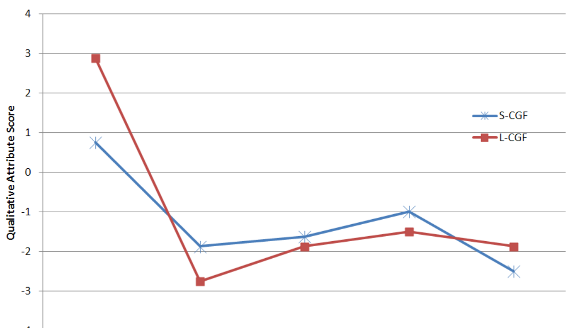


Fig. 11. Illustrations of the two *missile-fight* initial conditions used in the 2^{nd} HIL experiment

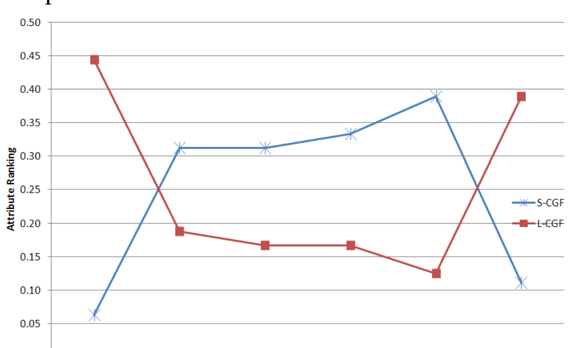
the same type of qualitative assessments by the trainee pilots shown in Fig. 7. In comparison, the overall opinion of the CGFs by the veteran pilots appears to be lower than that of the trainee pilots.



	Unpredictable	Intelligent	Skilful	Challenging	Adaptive
S-CGF	0.75	-1.875	-1.625	-1	-2.5
L-CGF	2.875	-2.75	-1.875	-1.5	-1.875

Fig. 12. Comparisons of the qualitative assessments of the L-CGF and the S-CGF by the veteran pilots using the trial questionnaire.

From Fig. 13, the qualitative assessments of the CGFs by the veteran pilots using the session questionnaires reveal a similar opinion shown in Fig. 12. It is in similar contrast to that of the trainee pilots in their direct comparison of the CGFs shown in Fig. 8. In this direct comparison, the L-CGF is perceived to be more predictable, less intelligent, less skillful, less challenging, less adaptive but more aggressive than the S-CGF by all the veteran pilots.



	Predictable	Intelligent	Skilful	Challenging	Adaptive	Aggressive
S-CGF	0.06	0.31	0.31	0.33	0.39	0.11
L-CGF	0.44	0.19	0.17	0.17	0.13	0.39

Fig. 13. Direct comparisons of the L-CGF and the S-CGF by the veteran pilots using the session questionnaire.

Quantitative Data: From Fig. 14, the L-CGF is shown losing to veteran pilot P2 for the 1^{st} four trials. Subsequently from the 5^{th} trial onwards, the L-CGF draws more with the veteran pilot P2. Notably, L-CGF scored a sortie win against veteran pilot P2 at the 6^{th} trial. This HIL experiment continued with veteran pilot P1 at the 7^{th} and 8^{th} trial. Using the knowledge learned from the 1-v-1 air combat with veteran pilot P2, L-CGF is able to draw for quite a number of sorties with veteran pilot P1. Veteran pilot P3 flew with the L-CGF for a single trial of 15 sorties at the 9^{th} trial. The L-CGF was also able to achieve draws with veteran pilot P3 for quite a number of sorties.

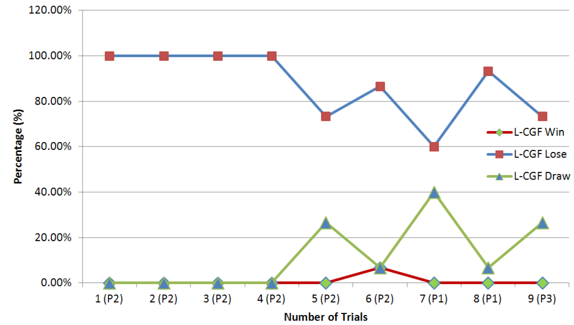


Fig. 14. Plots of the quantitative data from 1-v-1 dogfight between the L-CGF and the veteran pilots

From Fig. 15, the S-CGF is shown losing to the veteran pilot P2 consistently for the 1^{st} four trials. Imprecise timing of the maneuvers and using different strategies at the 5^{th} and 6^{th} trial allow the non-adaptive S-CGF to score a number of draws with veteran pilot P2. The time required by the veteran pilot P1 to adapt to the S-CGF also resulted in a number of draws with the S-CGF. The number of draws is reduced at his 2^{nd} trial with the S-CGF. At the 9^{th} trial, the need for the veteran pilot P3 to adapt to the S-CGF resulted in more draws between them.

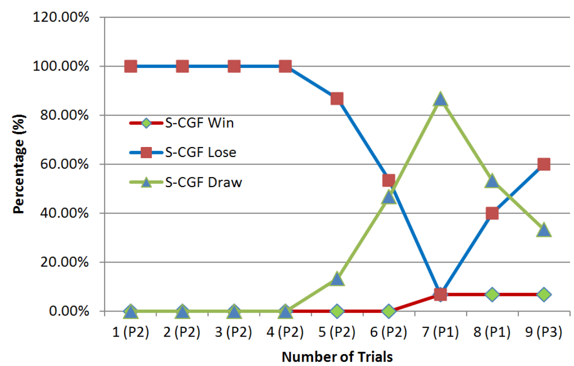


Fig. 15. Plots of the quantitative data from 1-v-1 dogfight between the S-CGF and the veteran pilots

D. Discussions

Two HIL experiments are conducted to study the performance of the adaptive CGF engaging in simulated 1-v-1 air combat against two groups of human pilots. The first group of human pilots comprising of two trainee pilots went through the HIL experiment using a *gun-fight* scenario against an adaptive CGF (L-CGF) and a non-adaptive CGF (S-CGF). Using blind qualitative assessment of the CGFs, both trainee pilots were able to match the desirable attributes to the L-CGF. From Fig. 8, the qualitative differences between the CGFs become more apparent when both of them are compared directly.

The quantitative results in Fig. 9 show the L-CGF losing to the trainee pilots due to the lack of desirable counter-strategies at the earlier trials. Over time, L-CGF is shown to be able to win more against the trainee pilots as it discards the bad strategies and learn the effective strategies. However, being better at adapting to the changing conditions, the trainee pilots are able to identify the more effective strategies against the L-CGFs at the later trials.

Three veteran combat pilots participated in the 2^{nd} HIL experiments. At their request, the initial conditions were changed to reflect a *missile-fight* scenario rather than a *gun-fight* scenario. Therefore, two *missile-fight* scenarios were used for the 2^{nd} HIL experiments. Using the new scenarios, the veteran pilots were able to see the L-CGF changing to the defensive strategies after a number of sorties. However, it did not survive long enough to a similar kind of outcome observed in Fig. 9.

In contrast to the 1^{st} HIL experiment, the defensive strategies are more effective against the human pilots than the offensive strategies in the chosen *missile-fight* scenarios. From Fig. 14, the L-CGF is still able to pick up some defensive strategies to manage a number of draws with the veteran pilots at the later trials. On the whole, the CGFs have greater difficulties winning against the veteran pilots. As a result, unlike the trainee pilots, the veteran pilots are not able to match the desirable qualitative qualities to the adaptive CGF. In this case, the adaptation of the L-CGF caused it to perform worse for most of the trials in the 2^{nd} HIL experiment.

VII. Conclusion

This work was conducted in collaboration with a multi-national simulator manufacturer to assess the performance of the adaptive CGF driven by TD-FALCON for training human pilots on 1-v-1 air combat. Two HIL experiments were conducted using commercial-grade simulation platform. Two trainee pilots and three veteran pilots were invited to assess the performance of the adaptive CGF (L-CGF) with respect to a non-adaptive doctrine-driven CGF (S-CGF). The identity of the CGFs is hidden from the pilots. Trial and session questionnaires are used to assess the CGFs qualitatively.

Wartime casualties are often the victim of the opponents with the unexpected maneuvers. Therefore, the pilots need to be trained to respond well to such opponents. An adaptive CGF is more suitable for the modeling of such opponent than a non-adaptive doctrine-driven CGF. This is evident from the initial challenges the adaptive CGF posed to the trainee pilots and also to veteran pilots at different times. Unlike the non-adaptive doctrine-driven CGF, this element of surprise is ever present in the adaptive CGF. The matching of the desirable attributes to the adaptive CGF by the trainee pilots is another sign of its value in the training simulator. On the other hand, the veteran pilots have helped us to be aware that the adaptive CGF is not adapting well for the *missile-fight* scenarios.

Though the adaptive CGF is shown to be better than the non-adaptive doctrine-driven CGF, it is still far from being able to match up to its human counterparts. Firstly, the adaptive CGF needs to improve on its adaptation rate with respect to the human pilots. Secondly, the adaptive CGF need to capable of adapting to fundamentally different strategies such as choosing between defensive and offensive strategies correctly. Beyond that, the execution of the HIL experiments can also be improved further. These are all the hot issues that should be adequately addressed for an adaptive CGF to be effective against a human opponent. In addition, there are plans to extend the use of the learning system to learn team coordination using a cognitive architecture with higher cognitive functions [4].

Acknowledgment

This work is supported by the DSO National Laboratories under research grant DSOCL11258. This project was conducted in close collaboration with Khee-Yin How and his team at DSO National Laboratories, Seng-Beng Ho and his team at Temasek Laboratories @ NUS, Adrian Yeo and his team at CAE Singapore (S.E.A.) Pte. Ltd. and Sylvain Caron and his team at CAE (Montreal) Inc. This work is dedicated to the memories of Wallace Peel, an CAF veteran combat pilot from Sylvain's team who contributed invaluably to this work.

References

[1] D. E. Forsythe and B. G. Buchanan, "Knowledge acquisition for expert systems: some pitfalls and suggestions," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 3, pp. 435 –442, May/Jun 1989.

[2] R. H. Ashton, "Combining the judgments of experts: How many and which ones?" *Organizational Behavior and Human Decision Processes*, vol. 38, pp. 405–414, 1986.

[3] D. Lenat, "The voice of the turtle: Whatever happened to AI?" *AI Magazine*, vol. 29, no. 2, pp. 11–22, 2008.

[4] G.-W. Ng, Y.-S. Tan, L.-N. Teow, K.-H. Ng, K.-H. Tan, and R.-Z. Chan, "A cognitive architecture for knowledge exploitation," in *Proceedings of the 3^{rd} Conference on Artificial General Intelligence*, 2010.

[5] F.-H. Hsu, "IBM's Deep Blue Chess grandmaster chips," *Micro, IEEE*, vol. 19, no. 2, pp. 70 –81, Mar/Apr 1999.

[6] D. Ferruci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, "Building Watson: An Overview of the DeepQA Project," *AI Magazine*, vol. 31, no. 3, pp. 59–79, 2010.

[7] H. H. Bell and W. L. Waag, "Evaluating the effectiveness of flight simulators for training combat skills: A review," *International Journal of Aviation Psychology*, vol. 8, no. 3, pp. 223–242, 1998.

[8] T.-H. Teng, A.-H. Tan, Y.-S. Tan, and A. Yeo, "Self-organizing Neural Networks for Learning Air Combat Maneuvers," in *Proceedings of the IJCNN (accepted)*, June 2012.

[9] A.-H. Tan, N. Lu, and X. Dan, "Integrating Temporal Difference Methods and Self-Organizing Neural Networks for Reinforcement Learning with Delayed Evaluative Feedback," *IEEE Transactions on Neural Networks*, vol. 19, no. 2, pp. 230–244, February 2008.

[10] P. A. Luker and B. Schmidt, Eds., *Knowledge-Based Simulation: methodology and application*. Springer-Verlag, 1990.

[11] N. R. Nielsen, "Applicability of AI Techniques to Simulation Models." in *Proceedings of the Simulation Series*, vol. 18, no. 4, 1987, pp. 35–40.

[12] K. Krishnakumar and S. Sawhney, "Simulator based adaptive helicopter training using neural networks," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics - Decision Aiding for Complex Systems*, oct 1991, pp. 1511–1516 vol. 3.

[13] J. Kaneshige and K. Krishnakumar, "Artificial immune system approach for air combat maneuvering," in *Proceedings of SPIE*, 2007, pp. 1–12.

[14] C. M. Oresky, D. B. Lenat, A. Clarkson, and S. H. Kaisler, *Knowledge-Based Simulation: Methodology and Application*. Springer-Verlag, 1991, ch. Strategic Automatic Discovery System (STRADS).

[15] H. K. G. Fernlund, A. J. Gonzalez, M. Georgiopoulos, and R. F. DeMara, "Learning tactical human behavior through observation of human performance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 1, pp. 128–140, Feb. 2006.

[16] B. Wrede, K. Rohlfing, M. Hanheide, and G. Sagerer, "Towards learning by interacting," in *Creating Brain-Like Intelligence*, ser. Lecture Notes in Computer Science, 2009, vol. 5436, pp. 139–150.

[17] A. Steinfeld, S. R. Bennett, K. Cunningham, M. Lahut, P. A. Quinones, D. Wexler, D. Siewiorek, J. Hayes, P. Cohen, J. Fitzgerald, O. Hansson, M. Pool, and M. Drummond, "Evaluation of an integrated multi-task machine learning system with humans in the loop," in *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, 2007, pp. 182–188.

[18] F. Sempé, M. Nguyen-Duc, S. Boissau, A. Boucher, and A. Drogoul, "An artificial maieutic approach for eliciting experts' knowledge in multi-agent simulations," *Lecture Notes in Computer Science*, vol. 3891, pp. 75–87, 2006.

[19] A.-H. Tan, "FALCON: A Fusion Architecture for Learning, Cognition, and Navigation," in *Proceedings of the IJCNN*, 2004, pp. 3297–3302.

[20] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.

[21] A.-H. Tan, "Direct Code Access in Self-Organizing Neural Networks for Reinforcement Learning," in *Proceedings of the IJCAI*, January 2007, pp. 1071–1076.

[22] A. Pérez-Uribe, "Structure-adaptable digital neural networks," Ph.D. dissertation, Swiss Federal Institute of Technology-Lausanne, 2002.

[23] M. D. Ardema and N. Rajan, "An approach to three-dimensional aircraft pursuit-evasion," *Computers & Mathematics with Applications*, vol. 13, no. 1-3, pp. 97–110, 1987.

[24] D. Gopher, M. Well, and T. Bareket, "Transfer of skill from a computer game trainer to flight," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 36, no. 3, pp. 387–405, 1994.

[25] R. Pausch, T. Crea, and M. Conway, "A literature survey for virtual environments: military flight simulator visual systems and simulator sickness," *Presence: Teleoperators and Virtual Environment*, vol. 1, pp. 334–363, 1992.