Nigerian Journal of Technology (NIJOTECH)
Vol. 42, No. 1, March, 2023, pp.144 - 151
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://doi.org/10.4314/njt.v42i1.18

DEEP REINFORCEMENT LEARNING FOR AIRCRAFT LONGITUDINAL CONTROL AUGMENTATION SYSTEM

A. O. Adetifa[1], P.P. Okonkwo[2], B. B. Muhammed[3], and D. A. Udekwe[4,*]

[1,2,3,4] Faculty of Air Engineering, Air Force Institute of Technology, Nigeria

*corresponding author (Phone Number: +234-706-764-7898. Email: nickudekwe@gmail.com)

Article history: Received 25 January, 2023. Revised 06 April, 2023. Accepted 11 April, 2023

Abstract
Control augmentation systems (CAS) are conventionally built with classical controllers which have the following drawbacks: dependence on domain specific knowledge for tuning and limited self-learning capability. Consequently, these drawbacks lead to sub-optimal aircraft stability and performance when exposed to time varying disturbances. Hence, to curb the stated problems; this paper proposes the development of a deep reinforcement learning (DRL) pitch-rate CAS (qCAS), aimed at guaranteeing adaptive stability, pitch-rate control tracking and disturbance rejection across the longitudinal dynamics of an aircraft. This stated aim was actualized by developing a CAS with a deep deterministic policy gradient (DDPG) agent. Subsequently, this proposed method was compared with two classical qCAS methods (a developed PID-aCAS and a benchmark PI-qCAS obtained from literature). The results show that the developed DDPG-qCAS method outperformed the classical methods in peak overshoot, referemce command tracking and disturbance rejection as well as mean absolute error (MSE) and mean steady state error (MSSE). Hence, it can be inferred that it is important to apply artificially intelligent controllers to the flight control systems of aircraft in order to achieve superior time response, control command tracking accuracy and disturbance rejection.

Keywords: Stability, Reference tracking, Disturbance rejection, Self-learning, Control augmentation, Longitudinal dynamics, Policy gradient.

1.0 INTRODUCTION

Guaranteeing stability is a crucial task in aircraft design as the performance of an aircraft depends on the ease with which it can be controlled and stabilized even in the presence of disturbances [1]. Aircraft disturbances refer to atmospheric phenomena such as wind gusts, wind gradients or turbulent air as well as unwanted control actions effected by the pilot [2]. An airplane must be sufficiently stable such that the pilot does not become fatigued by constantly having to manually trim the airplane following the occurrence of external disturbances [3].

It is important to note that, although airplanes with little or no inherent aerodynamic stability can be flown, such airplanes are unsafe to fly unless they are equipped with a form of feedback control system. These systems are usually classified into stability and control augmentation systems (SAS and CAS) [4]. A SAS is an electromechanical device for providing artificial stability in the presence of external disturbances [2] while a CAS provides not just stability but is designed to undertake specific control functions to reduce the workload of the pilot. These functions are usually in the form of tracking control of one or more of the aircraft's motion parameters [5].

Owing to the importance of these systems, an increasing amount of research has been directed towards the design of CAS for modern aircraft using various control methods. [6] presented the design of an adaptive control augmentation system for an F-16 model. This was aimed at increasing robustness in the presence of parametric uncertainties. In the work of [7] a proportional integral (PI) controller applied to the lateral and longitudinal dynamics of an F-16 aircraft for augmenting stability was studied. Similarly, the design of a stability augmentation control system for a turbojet aircraft was also considered by [3] using classical control. Subsequently, [8] investigated the stability augmentation problem of a flexible aircraft

for the purpose of improving the aircraft's dynamic stability.

Meanwhile, other research works have attempted to determine the relevance of optimal, robust and intelligent control schemes for assessing either the stabilization or command tracking properties of these systems. The paper by [9] presented the development of a longitudinal CAS for an aircraft based on LQR control tuned by a hybrid genetic algorithm (GA). [10] reported the design of stability augmentation systems (SAS) for small aircraft using the principles of robust control. Similarly, [11] presented the design of a fault-tolerant CAS for an F-16 aircraft with asymmetric elevator failures using $H\infty$ control. In another interesting study by [12] a robust pitch rate CAS for an F-16 aircraft was designed by $H\infty$ control as well. Finally, [13] presented improved reinforcement learning control for a quadcopter stability augmentation system using proximal policy optimization.

The remaining parts of this manuscript are organized in the following order; Section 2.0 entails the mathematical modelling of the longitudinal dynamics of an F-16 aircraft. Section 2.1 assesses the stability and control solution feasibility of the plant before integrating a controller. Section 2.2 presents the design of the DDPG-qCAS while in section 3.0, a report of the significant results obtained from the simulatory experiments is given. Section 4.0 summarizes the paper and provides recommendation for future works.

2.0 MODELLING THE LONGITUDINAL DYNAMICS OF AN F-16 AIRCRAFT

The pictorial illustration of an F-16 aircraft is shown in Figure 1. From a mathematical standpoint, the aircraft can be modelled by giving a description of the forces and moments acting on it.

According to [11] these forces and moments can be determined by means of wind tunnel tests carried out on a model of the aircraft and expressed as:

$$X = C_x\bar{q}S; \quad Y = C_y\bar{q}S; \quad Z = C_z\bar{q}S; \quad L = C_l\bar{q}Sb;$$
$$M = C_m\bar{q}S\bar{c}; \quad N = C_n\bar{q}Sb \tag{1}$$

Where $S, \bar{c}, \bar{q}, b$ represent wing area, wing chord, dynamic pressure, and wing span respectively. Furthermore, the coefficients of the dimensionless aerodynamic forces (C_x,C_y,C_z) and moments (C_l,C_m,C_n) [11].

The decoupled equations of pure longitudinal motion can therefore be given as follows (assuming no thrust-vectoring) [14], [15].

$$\dot{V} = \frac{\bar{q}S\bar{c}}{2mV}\left[C_{xq}(\alpha)\cos(\alpha) + C_{zq}(\alpha)sin\alpha\right] - gsin(\theta - \alpha) + \frac{\bar{q}S}{m}[C_x(\alpha,\delta_e)cos\alpha + C_z(\alpha,\delta_e)sin\alpha] + \frac{T}{m}\cos(\alpha) \tag{2}$$

$$\dot{\alpha} = q\left[1 + \frac{\bar{q}S\bar{c}}{2mV^2}\left(C_{zq}(\alpha)cos\alpha - C_{xq}sin\alpha\right] + \frac{\bar{q}S}{mV}[C_z(\alpha,\delta_e)cos\alpha - C_x(\alpha,\delta_e)sin\alpha + \frac{g}{V}\cos(\theta - \alpha) - \frac{T}{mV}\sin(\alpha) \tag{3}$$

$$\dot{\theta} = q \tag{4}$$

$$\dot{q} = \frac{\bar{q}S\bar{c}q}{2I_yV}\left[\bar{c}C_{mq}(\alpha) + \Delta C_{zq}(\alpha)\right] + \frac{\bar{q}S\bar{c}}{I_y}\left[C_m(\alpha,\delta_e) + \frac{\Delta}{\bar{c}}C_z(\alpha,\delta_e)\right] \tag{5}$$

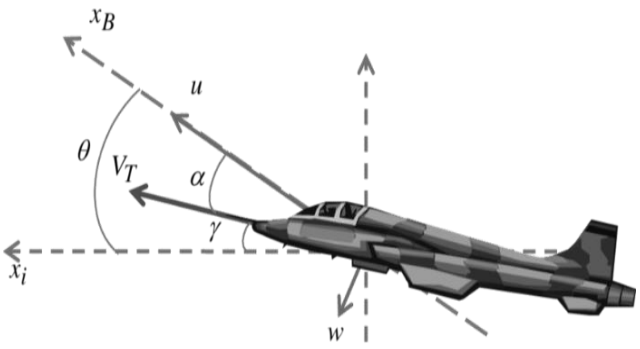


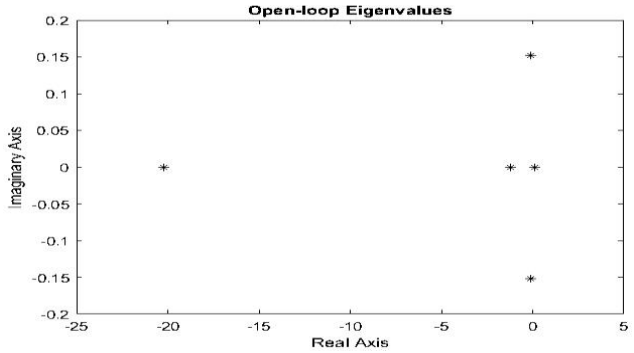
Figure 1: An F-16 aircraft showing longitudinal motion parameters [16]

A linearized variant of the aircraft's longitudinal dynamics can be obtained for steady wing-level flight using Taylor's series expansion and neglecting higher order terms according to [17].

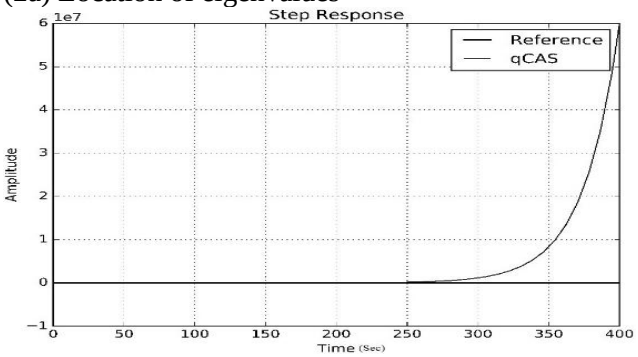
$$\begin{bmatrix} \triangle\dot{V} \\ \triangle\dot{\alpha} \\ \triangle\dot{\theta} \\ \triangle\dot{q} \end{bmatrix} = \begin{bmatrix} -0.022 & -1.395 & -9.828 & -0.672 \\ -0.002 & -0.582 & 0 & 0.908 \\ 0 & 0 & 0 & 1 \\ 3\times10^{-7} & 0.324 & 0 & -0.708 \end{bmatrix}\begin{bmatrix} V \\ \alpha \\ \theta \\ q \end{bmatrix} + \begin{bmatrix} -1.139 \\ -0.072 \\ 0 \\ -4.301 \end{bmatrix}[\delta_e] \tag{6}$$

$$\begin{bmatrix} q \\ \theta \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} V \\ \alpha \\ \theta \\ q \end{bmatrix} \tag{7}$$

2.1 Control Solution and Preliminary Stability Test



(2a) Location of eigenvalues



(2b) Open-loop step response

Figure 2: Stability of an F-16 aircraft along the longitudinal axis without a controller

A plot of the eigenvalues of the aircraft's longitudinal dynamics was used in determining the open-loop stability of the aircraft along its longitudinal axis. From the eigenvalue plot shown in 2(a), it is seen that the longitudinal dynamics of the aircraft are unstable and this is due to the positive root that exists at the right hand of the s- plane. To further illustrate this instability, the open-loop step response is given in 2(b) and it is shown that the system's response increases exponentially by the application of a bounded input (step signal). Additionally, it is reported that the longitudinal dynamics of an F-16 aircraft are controllable and observable because the controllability and observability matrices existed in full rank (N = 4). This observation necessitated the need to develop a controller that would guarantee stability; thereby the design of a CAS that alters the location of the system's eigenvalues to the left-hand side of the s-plane was undertaken.

2.2 PID-qCAS Design

A block diagram of the PID based qCAS developed in the Simulink environment is shown in Figure 3. The inner loop provides dynamic stability to the longitudinal motion of the aircraft while the outer loop is responsible for pitch rate tracking. After many iterations the feedback gain K_a was determined to be 0.08 while the parameter gains (K_p, K_i and K_d) of the PID controller were determined using the MATLAB PID tuner app in accordance with equation (8).

$$u(t) = Kp \times e(t) + K_i \int_0^t e(t)\,dt + K_d \frac{de(t)}{dt} \quad (8)$$

The response of the system while taking a unit step signal is shown in Figure 6 while the time response is given in Table 2.

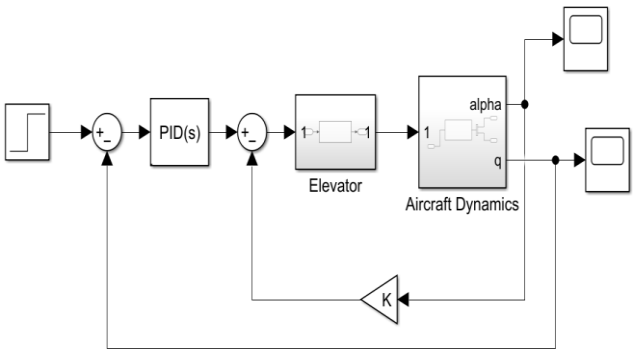


Figure 3: PID based pitch rate control augmentation system

2.3 DDPG-qCAS Design

Deep reinforcement learning entails the combination of reinforcement learning with deep learning for the purpose of building computationally intelligent algorithms. These algorithms are envisaged to be capable of interacting with a dynamic environment based on a predetermined criterion to optimize a long-term reward [18]. Deep reinforcement learning has been successfully implemented for traffic signal control [19] load balancing [20], vehicular motion control [21] and anomaly detection [20].

Deep deterministic policy gradient (DDPG) is an actor critic reinforcement learning method which was originally presented by [22] for solving continuous action tasks. In this study, a DDPG reinforcement learning algorithm was employed to train an agent which is expected to estimate a control command $u(t)$ shown in equation (9) based on its interactions with the environment through the states represented in (10).

$$\big(u(t)\big) = a_k = \pi(s_k) \quad (9)$$

$$S = \left(e(z), e(z) \times \frac{KT_s}{z-1}, e(z) \times \frac{K(z-1)}{T_s z}\right) \quad (10)$$

The steps needed for implementing a DDPG-qCAS are described as follows:

1. Define the environment: The environment contains the following components: the dynamic

model of the aircraft's longitudinal dynamics, the reward generating block, and the stopping conditions which provide Boolean answers: yes(1), or no(0) as well as the observations blocks. These can be described as the world or physical system which consists of all components excluding the reinforcement learning (RL) agent.

2. Define the reward function: This evaluates the performance of the agent at every time step by computing the effective value function. This was implemented by a quadratic cost function which is described in equation (11)

$$R_i = 10 - (10e(t)^2 + 5u(t)^2 + 10(OR(H1, H2))) \quad (11)$$

where $H1$ represents $y_i \leq 1.6$ and $H2$ represents $y_i \geq 2$; $e(t)$ denotes the predictive error, and $u(t)$ accounts for the control law in the time-domain.

The stopping condition block checks the OR logic operation. If the stated condition is met, then it is expected to return a Boolean value of 1, otherwise, it would return a value of 0.

3. Creation of the agent: A policy and a learning algorithm are the main components of the agent. DDPG agents depend on critic and actor networks to approximate the policy and value function. Furthermore, DDPG agents are also reliant on the target network and replay buffer (also known as experience replay) which is an activity done during training of the agent to possibly sample for several mini-batches. The actor network employs the use of a policy gradient algorithm for computing the plant control signal $u(t)$ and to improve exploration of the agent, Gaussian noise is factored in according equation (12):

$$a_i^t = N_t + \mu(S_i^t|\theta^\mu) \quad (12)$$

Where the μ refers to the actor parameter, given an observation state S_t θ^μ denotes the actor learned weights, and the variable N_t accounts for the Gaussian noise.

The central goal of training the agent is to ensure that maximum rewards are obtained in the long run. Hence, the cumulative reward can be computed using the expression below:

$$y_i = R_i + \gamma Q^{t+1}\left(S_i^{t+1}, \mu^{t+1}\left(S_i^{t+1}/\theta_\mu\right)\theta_{Q'}\right) \quad (13)$$

The critic component computes the long-term expected rewards of the agent based on its current state s and actions a and it is updated by minimizing a loss function (L) over a number of sample experiences n through the back propagation technique. This is achieved by optimizing the learning weights using gradient descent optimization.

$$L = \frac{1}{n}\sum_{i=1}^{n}[y_i - Q^t(S_i^t, a_i^t|\theta^Q)]^2 \quad (14)$$

The actor network is updated by sampling the policy gradient:

$$\nabla_{\theta^\mu} \approx \frac{1}{N}\sum_i \nabla_a Q(s_i, a_i|\theta^Q)|_{s_i=s_i^t, a_i=\mu(s_i^t)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|s_i^t \quad (15)$$

Additionally, the target actor and critic network are updated according to equations (16) and (17). These networks are simply time delayed replicas of the main networks created for the purpose improving the stability of the optimization process.

$$\theta^{Q'} \leftarrow (1-\tau)\theta^{Q'}\ \tau\theta^Q \quad (16)$$

$$\theta^{\mu'} \leftarrow (1-\tau)\theta^{\mu'}\ \tau\theta^\mu \quad (17)$$

Where the variable τ denotes the smoothing factor for the critic and actor networks.

Figure 4 shows the architectural representation of the novel DDPG agent used in implementing the DDPG-qCAS controller.

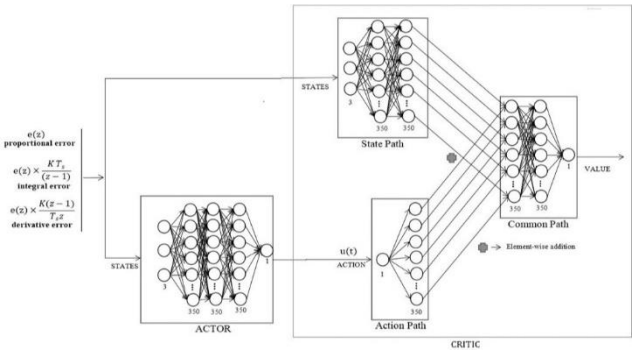


Figure 4: Architectural representation of DDPG agent used in implementing DDPG-qCAS

The DDPG agent used in implementing the qCAS consists of actor and critic networks containing 350 network units in each of the hidden or bottleneck layers. A rectified linear unit (ReLU) for processing the network pre-activation features within the bottleneck and the outer network layers was factored. The overall closed-loop system of the developed qCAS is described with a block diagram consisting of

the already described DDPG agent integrated with the state-space representation of an F-16 longitudinal dynamics. This is shown in Figure 5. The following experimental parameters shown in Table 1 were specified for the purpose of training the agent and the plot of the learning curve over 1420 episodes is shown in Figure 6.

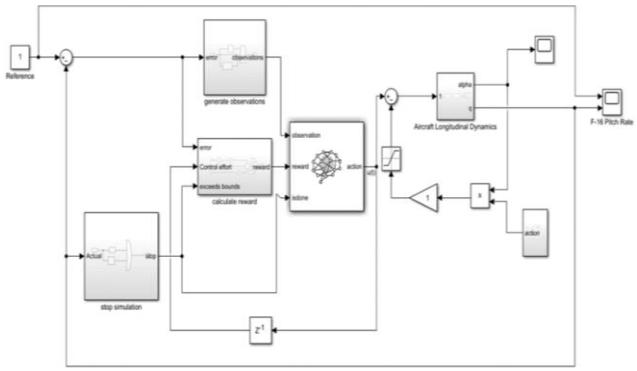


Figure 5: A pictorial illustration of the DDPG-qCAS showing the integration of the DRL controller and the longitudinal model of the aircraft.

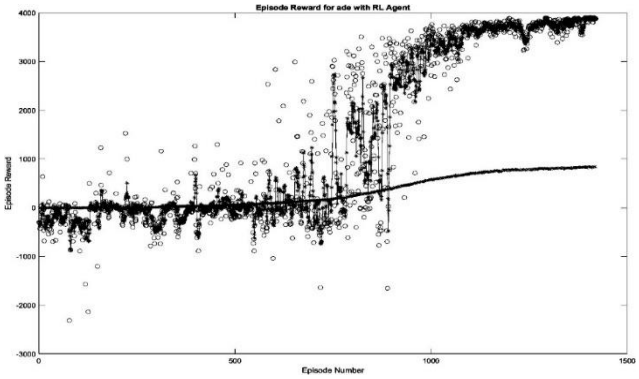


Figure 6: A graphical illustration of the episodic rewards generated during training the DRL agent for 1420 episodes

Table 1: DRL experimental parameters

Parameter	Value
Maximum number of episodes	1420
Actor learning rate	1×10^{-5}
Critic learning rate	1×10^{-4}
Minimum steps per episode	40
Discount Facto	0.99

Experimental setup: Training the DDPG agent for 1420 episodes was performed in the Simulink environment using the following hardware resource for about 420 minutes: a Dell Latitude Laptop with 16GB RAM, 512GB SSD hard-disk, Intel(R) processor Core(TM) i7 CPU @ 2.6GHz.

3.0 RESULTS AND DISCUSSION

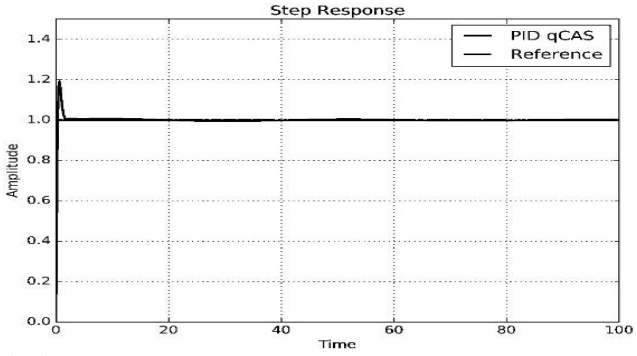
This section provides an evaluation of the DDPG-qCAS based on the following:

1. Error metrics: integral square error (ISE), mean steady-state error (MSSE), and mean absolute error (MAE).
2. Time responses: settling time (t_S), peak overshoot (M_p), and rise time (t_r).

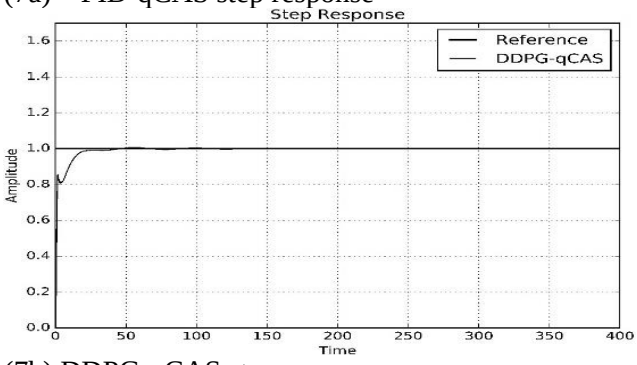
Furthermore, the command tracking and disturbance rejection performance of these systems are discussed.

3.1 Time Response and Predictive Errors

The response of the DDPG-qCAS to a unit step signal is shown in Figure 7 and the time response and error metrics are given in Table 2.



(7a) PID-qCAS step response



(7b) DDPG-qCAS step response

Figure 7: Step response of PID-qCAS and DDPG-qCAS

Table 2: Time response and predictive error comparison between DDPG-qCAS and PID-qCAS

Method	$t_r(s)$	$t_s(s)$	M_p	MAE	ISE	MSSE
PID-qCAS	0.2386	1.4169	18.9721	0.1018	0.1196	0.0892
DDPG-qCAS	9.4061	17.9802	0.6590	0.0100	0.8861	0.0138

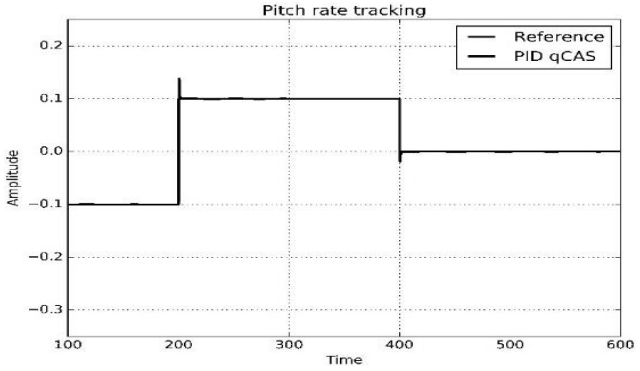
3.2 Pitch-Rate Command Tracking

A reference pitch rate command signal according to [11] was developed in the MATLAB environment using the signal builder tool. This signal

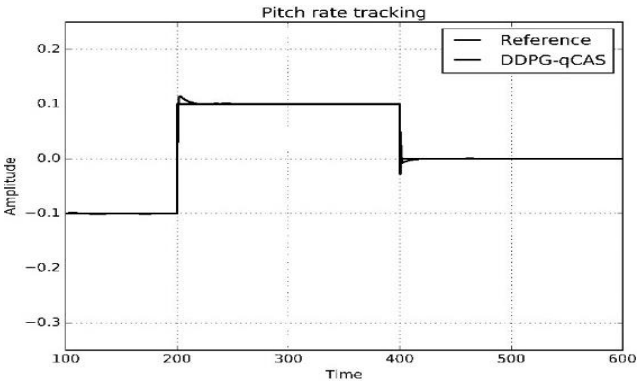
hypothetically represents a desired pitch rate of an F-16 aircraft at a particular flight condition for a period of 600 seconds. Simulation carried out in the MATLAB environment shows that it was sufficient to track a reference pitch rate signal with minimum error. A summary of this performance based on mean absolute error (MAE) is shown in Table 3.

Table 3: Comparison between command tracking performance of PID-qCAS and DDPG-qCAS.

Techniques	Command Tracking MAE
PID-qCAS	0.0239
DDPG-qCAS	**0.0069**



(8a) PID-qCAS command tracking



(8b) DDPG-qCAS command tracking

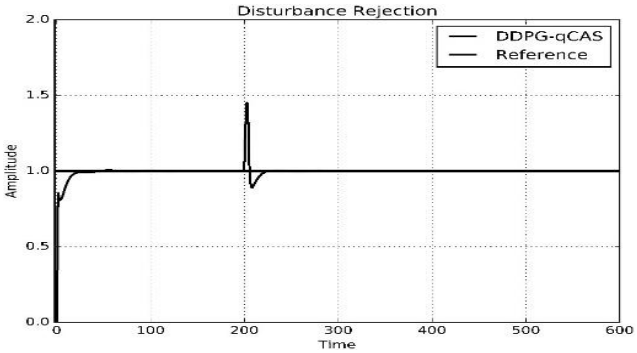
Figure 8: Command tracking performance of an F-16 aircraft integrated with either or DDPG-qCAS

3.3 Disturbance Rejection

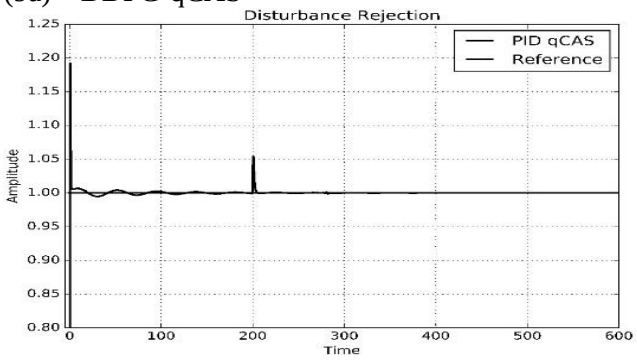
Disturbance rejection assesses the robustness of the proposed systems when exposed to external disturbances. In this test, the aircraft is expected to maintain zero pitch rate deflection when exposed to an enormous pitching moment disturbance of 30kN occurring at the 200th second while flying at trim. Figure 8 shows that the developed DDPG-qCAS was adequately robust to this disturbance and as a result was sufficiently able to restore the aircraft back to trim. The quantitative assessment of this method for disturbance rejection is presented in Table 4.

Table 4: Comparison between disturbance rejection performance of PID-qCAS and DDPG-qCAS.

Technique	Disturbance Rejection MAE
PID-qCAS	0.0649
DDPG-qCAS	**0.0161**



(9a) DDPG-qCAS



(9b) PID-qCAS

Figure 9: Disturbance rejection performance of an F-16 aircraft integrated with either PID-qCAS or DDPG-qCAS

To further assess the goodness of the novel method, the DDPG-qCAS was compared with a benchmark method obtained from literature and this comparison is discussed in the subsection below.

3.4 Benchmark Validation: Comparing the novel DDPG-qCAS with the PI-qCAS presented by [21]

The DDPG-qCAS was compared with the PI-qCAS developed by [23]. The report of the comparison is presented in Table 5. From the result table, it can be observed that the developed DDPG-qCAS achieved a better peak overshoot and peak value of 0.6590%, and 1.0066, respectively, compared to the PI-qCAS which obtained 11.80%, and 1.1179. Nonetheless with the drawback of a poorer rise time and settling time. This superior performance however can be attributed to the use of a deep reinforcement learning model that employs neural network function approximators and an updated optimal learning policy whose central goal is to generate an optimal control action that maximizes a long-term reward (minimization of steady state error).

Table 5: Performance comparison of the novel DDPG-qCAS with the PI-qCAS [23] analyzed on an F-16 aircraft

Method	Rise Time	Settling Time	Peak Overshoot	Peak Value
DDPG-qCAS	9.4061	17.9802	0.6590	1.0066
PI-qCAS [21]	0.15	1.52	11.80	1.1179

4.0 CONCLUSION

The longitudinal stability, command tracking as well as disturbance rejection problems of an aircraft were evaluated by developing a pitch rate control augmentation system (qCAS) based on deep reinforcement learning. This developed control augmentation methods denoted as DDPG-qCAS represents a pitch-rate control augmentation systems for an aircraft developed by a deep deterministic policy gradient (DDPG) agent. The evaluation of this method was performed by comparing with two classical methods based on the time response, minimization of the errors, command tracking accuracy and disturbance rejection. The results show that the proposed DDPD-qCAS method yielded suitable performance and surpassed the classical methods in peak overshoot, reference command tracking and disturbance rejection as well as mean absolute error (MSE) and mean steady state error (MSSE). This superior performance can be attributed to the use of neural networks based on non-linear function approximators and an updated learning algorithm. Hence, it can be inferred that it is important to apply artificially intelligent controllers to the flight control systems of aircraft in order to achieve superior command tracking and disturbance rejection performance as well as minimum error metrics.

Future research can explore other deep reinforcement learning algorithms such as trust-region proximal optimization, asynchronous advantage actor-critic, and other behavioural cloning techniques compared with the developed method.

REFERENCES

[1] McLean, D. "Automatic flight control systems(Book)," *Englewood Cliffs, NJ, Prentice Hall, 1990, 606*, 1990.

[2] Nelson, R. C. *Flight stability and automatic control*, vol. 2. WCB/McGraw Hill New York, 1998.

[3] Putro, I. E., and Duhri, R. A. "Longitudinal stability augmentation control for turbojet UAV based on linear quadratic regulator (LQR) approach," in *AIP Conference Proceedings*, 2020, vol. 2226, no. 1, p. 20013.

[4] Jang, J. and Tomlin, C. "Longitudinal stability augmentation system design for the DragonFly UAV using a single GPS receiver," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2003, p. 5592.

[5] Stevens, Lewis, and Johnson, "Aircraft dynamics and classical control design," *Aircr. Control Simul.*, pp. 254–382, 2003.

[6] Bhardwaj, P., Akkinapalli, V. S., Zhang, J., Saboo, S., and Holzapfel, F "Adaptive augmentation of incremental nonlinear dynamic inversion controller for an extended f-16 model," in *AIAA Scitech 2019 Forum*, 2019, p. 1923.

[7] Huang, W., Zhao, Y., Ye, Y., and Xie, W. "State feedback control for stabilization of the ball and plate system," in *Chinese Control Conference, CCC*, 2019, vol. 2019-July, pp. 687–690, doi: 10.23919/ChiCC.2019.8866355.

[8] Barbosa, G. C., Bertolin, R. M., Paulino, J. A., Silvestre, F., and A. B. Guimarães Neto, "Design and Flight Test of a Stability Augmentation System for a Flexible Aircraft," in *AIAA Scitech 2019 Forum*, 2019, p. 365.

[9] Yuxing, Z., Jie, X., and Zhikuo, C. "LQR Controller for Longitudinal Control Augmentation System of Aircraft Based on Hybrid Genetic Algorithm," in *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*, 2018, pp. 1–4.

[10] Barufaldi, G. N., Victor Jr, M. H., Pereira, H. W. R., and da Silva, R. G. A. "A Robust Longitudinal Stability Augmentation System for Small Aircraft Under Parametric Uncertainty," in *Congresso Brasileiro de Automática-CBA*, 2020, vol. 2, no. 1.

[11] Gümüşboğa, I., and İftar, A. "Fault-tolerant pitch-rate control augmentation system design for asymmetric elevator failures in a combat plane," *Kybernetika*, vol. 56, no. 4, pp. 767–793, 2020.

[12] Gumusboga, I., "Robust Pitch Rate Control Augmentation System for an F-16 Aircraft," *J. Aeronaut. Sp. Technol.*, vol. 14, no. 2, pp. 243–249, 2021.

[13] Wu, H., Ye, H., Xue, W., and Yang, X., "Improved Reinforcement Learning Using Stability Augmentation With Application to Quadrotor Attitude Control," *IEEE Access*, vol. 10, pp. 67590–67604, 2022.

[14] Vo, H., and Seshagiri, S. "Robust control of F-16 lateral dynamics," in *2008 34th Annual*

Conference of IEEE Industrial Electronics, 2008, pp. 343–348.

[15] Promtun, E. and Seshagiri, S. "Sliding mode control of pitch-rate of an f-16 aircraft," *IFAC Proc. Vol.*, vol. 41, no. 2, pp. 1099–1104, 2008.

[16] Nguyen, V. H., & Tran, T. T. (2020). A novel hybrid robust control design method for f-16 aircraft longitudinal dynamics. *Mathematical Problems in Engineering*, *2020*, 1-10.

[17] Gümüşboğa, I and İftar, A. "Aircraft trim analysis by particle swarm optimization," *J. Aeronaut. Sp. Technol.*, vol. 12, no. 2, pp. 185–196, 2019.

[18] Sutton, R. S., and Barto, A. G. Reinforcement learning: An introduction. *Robotica*, *17*(2), 229-235; 2018.

[19] Kassem, A. Haddad, H., and Albitar, C. "Commparison Between Different Methods of Control of Ball and Plate System with 6DOF Stewart Platform," in *IFAC-PapersOnLine*, 2015, vol. 48, no. 11 11, pp. 47–52, doi: 10.1016/j.ifacol.2015.09.158.

[20] Annen, K. B., Stickler, D. M. and Woodroffe, J. "High energy density electric power for UAV/UGV propulsion using a Miniature IC Engine (mice)," *2nd AIAA "Unmanned Unlimited" Conf. Work. Exhib.*, 2003, doi: 10.2514/6.2003-6632.

[21] Buechel, M., and Knoll, A. "Deep reinforcement learning for predictive longitudinal control of automated vehicles," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2391–2397.

[22] Lillicrap, T. P., et al., "Continuous control with deep reinforcement learning," 2016, Accessed: Oct. 27, 2021. [Online]. Available: https://arxiv.org/abs/1509.02971.

[23] Gümüşboğa, I., and İftar, A. "Pitch-rate control augmentation system design for delayed measurements," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2019, pp. 940–945.