Journal of *Marine Science and Engineering*

MDPI

Article

An Efficient Multi-AUV Cooperative Navigation Method Based on Hierarchical Reinforcement Learning

Zixiao Zhu [1,2], Lichuan Zhang [1,2,*], Lu Liu [1,2], Dongwei Wu [3], Shuchang Bai [1,2], Ranzhen Ren [1,2] and Wenlong Geng [1,2]

[1] Research & Development Institute of Northwestern Polytechnical University, Shenzhen 518057, China; zixiaozhu@mail.nwpu.edu.cn (Z.Z.); liulu12201220@nwpu.edu.cn (L.L.); shuch-an_bai@mail.nwpu.edu.cn (S.B.); rrz@mail.nwpu.edu.cn (R.R.); gengwenlong@mail.nwpu.edu.cn (W.G.)
[2] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
[3] Shanghai Suixun Electronic Technology Co., Ltd., Shanghai 200438, China; wudw@mail.nwpu.edu.cn
* Correspondence: zlc@nwpu.edu.cn

Abstract: Positioning errors introduced by low-precision navigation devices can affect the overall accuracy of a positioning system. To address this issue, this paper proposes a master-slave multi-AUV collaborative navigation method based on hierarchical reinforcement learning. First, a collaborative navigation system is modeled as a discrete semi-Markov process with defined state and action sets and reward functions. Second, trajectory planning is performed using a hierarchical reinforcement learning-based approach combined with the polar Kalman filter to reduce the positioning error of slave AUVs, realizing collaborative navigation in multi-slave AUV scenarios. The proposed collaborative navigation method is analyzed and validated by simulation experiments in terms of the relative distance between the master and slave AUVs and the positioning error of a slave AUV. The research results show that the proposed method can not only successfully reduce the observation and positioning errors of slave AUVs in the collaborative navigation process but can also effectively maintain the relative measurement distance between the master and slave AUVs within an appropriate range.

Keywords: SMDP; AUV; cooperative navigation; hierarchical reinforcement learning; abstract action; Q-learning

Citation: Zhu, Z.; Zhang, L.; Liu, L.; Wu, D.; Bai, S.; Ren, R.; Geng, W. An Efficient Multi-AUV Cooperative Navigation Method Based on Hierarchical Reinforcement Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 1863. https://doi.org/10.3390/jmse11101863

Academic Editor: Rafael Morales

Received: 2 September 2023
Revised: 13 September 2023
Accepted: 16 September 2023
Published: 26 September 2023

1. Introduction

Automatic underwater vehicles (AUVs) have been widely used in marine science research and marine engineering and have become an essential tool for marine exploration and resource development [1,2]. As a collaborative-working tool, multi-AUV systems are efficient and flexible and can meet the requirements of complex tasks. In a multi-AUV cooperative navigation system, ensuring the accurate positioning of a slave AUV is crucial for the entire positioning system's performance. Currently, multi-AUV cooperative navigation systems can be roughly classified into parallel and master-slave systems [3–5]. In parallel systems, all AUVs are equipped with navigation tools with the same precision. Therefore, it is typically required to invest in expensive high-precision navigation equipment to enhance the positioning accuracy of the entire formation. In contrast, in the master-slave systems, AUVs are classified into master and slave AUVs based on the accuracy of their navigation equipment. This approach ensures the high positioning accuracy of the system while effectively controlling costs, which makes it the mainstream research direction.

The underwater navigation environment is complex, and navigation sensors are prone to anomalies. Therefore, enhancing the adaptability of cooperative navigation algorithms to outliers has become a research hotspot in recent years [6–10]. For instance, Bai [11] modeled the non-Gaussian noise introduced by sensor outliers as two advanced

types of mixed distributions. By introducing two Bernoulli random variables, these two mixed distributions are represented in a hierarchical Gaussian form, designing a robust Kalman filter based on the mixed distribution. Li et al. [12] proposed a robust multi-AUV cooperative navigation algorithm based on the student's extended Kalman filter, which can dynamically adapt to outliers in process noise and measurement noise. Xu et al. [13] developed an adaptive noise estimation algorithm based on a covariance matching method, which can adaptively estimate Gaussian and non-Gaussian measurement and processing noises. Considering the heavy-tailed measurement noise (heavy-tailed measurement noises) caused by non-Gaussian measurements in a cooperative positioning system, Sun et al. [14] designed an innovative maximum entropy variational difference filter based on the divided difference filter (DDF), which combines the advantages of DDF and the maximum entropy criterion, which enhances the robustness of the filter; the proposed filter was verified by lake experiments. Zhang et al. [15] introduced an AUV cluster network navigation accuracy analysis method based on the Fisher information matrix and demonstrated that using the information on the entire formation when estimating the position of the following AUV could improve the positioning accuracy of the following AUV. Chiarella D. [16] designed a multi-AUV framework, defining a hierarchical work order of AUVs to efficiently complete cooperative tasks, and proposed a gesture-based UHRI framework to optimize coordination and communication between AUVs while supporting the other communication methods.

This paper proposes a multi-AUV cooperative navigation method based on a hierarchical reinforcement learning-based algorithm for master-slave multi-AUV cooperative navigation systems. The proposed method divides the entire cooperative navigation task into two phases, the trajectory planning phase and the navigation processing phase. In the trajectory planning phase, a master-slave AUV cooperative navigation model is designed, and the concept of abstract actions is introduced, establishing the semi-Markov decision process (SMDP) model. Furthermore, a master AUV trajectory planning method based on the hierarchical model, which can reduce the observation and positioning errors of a slave AUV, is developed. In the navigation processing phase, the trajectory planning results are integrated with the Unscented Kalman Filter (UKF), an advanced extension of the Kalman filtering technique designed for state estimation in nonlinear systems. The UKF operates by selecting a representative set of sample points, termed sigma points, to approximate the mean and covariance of nonlinear functions, thereby circumventing the necessity for linearization. This combination realizes a comprehensive cooperative navigation method process. Subsequently, navigation simulation experiments are performed in two cooperative navigation scenarios with two slave AUV linear paths and three slave AUV serpentine search line paths to verify the effectiveness of the proposed method. By introducing a hierarchical structure and abstract actions, this study decomposes and hierarchically manages tasks in the main trajectory planning of multiple slave AUVs. This hierarchical structure can decrease the state space size and reduce the problem complexity, thus improving the efficiency and speed of problem-solving. Compared with traditional overall planning methods, the proposed hierarchical method can better handle the complexity of multiple AUV cooperative navigation tasks.

The rest of this paper is organized as follows. Section 2 analyzes the multi-AUV cooperative navigation models, including the kinematic model and cooperative navigation error model, and discusses the observability of a cooperative navigation system. Section 3 presents a multi-AUV cooperative navigation method based on hierarchical reinforcement learning, explaining related theories of Q-learning and hierarchical reinforcement learning, and introduces a trajectory planning method based on the hierarchical model. Section 4 simulates and verifies the proposed method, analyzes the simulation results, and explains the advantages of the proposed method compared to the existing methods. Finally, Section 5 summarizes the paper.

2. Multi-AUV Cooperative Navigation Model

Based on the number of master AUVs, multi-AUV cooperative navigation systems can be roughly divided into single- and multi-master multi-slave AUV cooperative navigation systems [17–20]. The former systems are more focused on due to the requirement for fewer high-precision navigation devices, making them more cost-effective than the latter systems. Therefore, this study adopts a single-master multi-slave AUV cooperative navigation method.

2.1. Cooperative Navigation Model

The multi-AUV system's cooperative navigation processing is illustrated in Figure 1. As shown in Figure 1, in the multi-AUV cooperative navigation system, AUVs exchange information through mutual communication for cooperative navigation, which improves the underwater navigation accuracy of AUVs. Typically, the master AUVs are often equipped with the integration of real-time data processing units for high precision positioning, while the slave AUVs are equipped with low-precision, low-cost navigation devices. Master and slave AUVs communicate and exchange information with each other through various communication devices, such as underwater modems [21]. For instance, in the single-master AUV and single-slave AUV scenarios, the AUVs communicate at a fixed interval. First, they measure the relative distance and relative azimuth angle between the AUVs using devices such as an Ultra-Short Base Line (USBL), which is an underwater acoustical positioning system employed to ascertain the position of subaqueous objects, such as submarines, autonomous underwater vehicles, or sensors [22]. Then, the master AUV sends information on its position to the slave AUV through the underwater modem. The slave AUV uses the received position information and the measured data on the distance and azimuth between the master and slave AUVs to estimate its current position, correcting the cumulative error introduced by dead reckoning (DR).

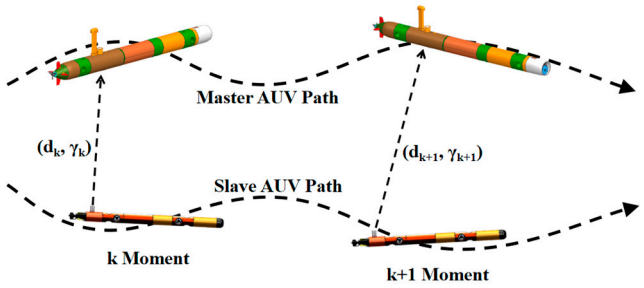

Figure 1. Illustration of the multi-AUV system cooperative navigation processing.

Before establishing the cooperative navigation model of AUVs, the first step involves defining the coordinate system. In this study, the North East Down (NED) coordinate system, which is a geocentric fixed coordinate system that has been widely used in geographic information systems and aerospace fields, is used. This system is selected because its definition is well-suited to address the problem studied in this work. In the NED coordinate system, the directions of the three axes are defined as directions pointing to the north of the Earth (*N*-axis), the east of the Earth (*E*-axis), and the center of the Earth (*D*-axis).

After defining the coordinate system, the motion model of a single AUV is established. Since the sailing depth of an AUV can be accurately measured by depth sensors, and the roll angle ϕ and pitch angle θ of the AUV slightly change during stable navigation, this study selects the eastward position x, northward position y, and heading angle ψ of an AUV to construct the system's state vector $\mathbf{x_k}$, which is expressed as $\mathbf{x_k} = [x_k \; y_k \; \Psi_k]^T$. An AUV's motion model is established in a two-dimensional plane [23,24]. In addition,

for simplicity of analysis, disturbances such as ocean currents are ignored. Based on the kinematic characteristics of a vessel, the following kinematic equations are established:

$$\begin{cases} x_{k+1} = x_k + TV_k \cos\psi_k \\ y_{k+1} = y_k + TV_k \sin\psi_k \\ \psi_{k+1} = \psi_k + T\omega_k \end{cases} \tag{1}$$

where x_{k+1} and y_{k+1} are the longitudinal and lateral coordinates of an AUV in the navigation coordinate system at a time $(k+1)$; ψ_{k+1} is the yaw angle of the AUV at a time $(k+1)$; V_k is the traveling speed of the AUV at a time k; ω_k is the yaw rate of the AUV at a time k; T is the sampling period of the AUV.

In this model, the input measured by a sensor $\mathbf{u}_k$ is expressed as follows:

$$\mathbf{u}_k = \begin{bmatrix} V_k \\ \omega_k \end{bmatrix} = \begin{bmatrix} V_{k,m} + \sigma_{V,k} \\ \omega_{k,m} + \sigma_{\psi,k} \end{bmatrix} = \mathbf{u}_{k,m} + \mathbf{w}_k \sigma_{\psi,k}, \tag{2}$$

where V_k is the velocity of an AUV at a time k; $V_{k,m}$ is the traveling speed of the AUV measured by the DVL at a time k; ω_k is the angular velocity of the AUV at a time k; $\omega_{k,m}$ is the yaw rate of the AUV measured by a gyroscope at a time k; $\sigma_{V,k}$ is the AUV's velocity measurement equipment error at a time k; $\sigma_{\psi,k}$ is the AUV's yaw angle rate measurement equipment error at a time k.

Combining Equations (1) and (2), the kinematic model of an AUV can be described by:

$$\mathbf{X}_{k+1} = f(\mathbf{X}_k, \mathbf{u}_k) = f(\mathbf{X}_k, \mathbf{u}_{k,m}, \mathbf{w}_k) = \mathbf{X}_k + \mathbf{\Psi}_k(\mathbf{u}_{k,m} + \mathbf{w}_k), \tag{3}$$

where $\mathbf{\Psi}_k = \begin{bmatrix} T\cos(\psi_k) & 0 \\ T\sin(\psi_k) & 0 \\ 0 & T \end{bmatrix}$ represents the nonlinear terms in the model.

Next, assume that Q is the system noise covariance matrix; then, it holds that:

$$\mathbf{Q}_k = E\left\{\mathbf{w}_k \mathbf{w}_k^T\right\} = \begin{bmatrix} \sigma_{V,k}^2 & 0 \\ 0 & \sigma_{\psi,k}^2 \end{bmatrix}. \tag{4}$$

For a simplified two-dimensional single-master AUV cooperative navigation system, the measured quantity is the distance between AUVs, which is calculated by:

$$d_{k+1} = \sqrt{(x_{k+1}^s - x_{k+1}^m)^2 + (y_{k+1}^s - y_{k+1}^m)^2} + \sigma_{d,k+1}, \tag{5}$$

where x_{k+1}^s and y_{k+1}^s represent the position coordinates of a slave AUV at a time $(k+1)$; x_{k+1}^m and y_{k+1}^m represent the position coordinates of a master AUV at a time $(k+1)$; $\sigma_{d,k+1}$ represents the distance measurement error of an acoustic measurement device at a time $(k+1)$.

After converting Equation (5) into the matrix form, the measurement equation is expressed by:

$$\mathbf{Z}_{k+1} = h(\mathbf{X}_{k+1}) + \mathbf{v}_{k+1}, \tag{6}$$

where $\mathbf{v}_{k+1}$ represents the measurement noise matrix.

Furthermore, assume that R is the covariance matrix of the system's measurement noise; then, it holds that:

$$\mathbf{R}_{k+1} = E\left\{\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\right\} = \left[\sigma_{d,k+1}^2\right]. \tag{7}$$

Through the above-presented analysis, a mathematical model for multi-AUV cooperative navigation has been established, providing a theoretical basis for the subsequent analysis.

2.2. Observability and Observation Error

To achieve multi-AUV cooperative navigation, the system should be observable, so it is necessary to analyze the system's observability. Furthermore, the analysis of the system's observability represents the theoretical basis for the cooperative navigation method proposed in this study. Therefore, this study analyzes the observability of a single-master AUV cooperative navigation system, conducts the observation error analysis, and derives the error propagation equation for a multi-AUV cooperative navigation system. Based on the proof given in [25], the observability of a multi-AUV cooperative navigation system can be proven. A multi-AUV cooperative navigation system is observable, and the relationship between the system's observability and the absolute difference in the azimuth angle of adjacent distance observations can be demonstrated.

The ultimate goal of cooperative navigation is to reduce cooperative positioning errors. In view of that, this study analyzes the observation error characteristics of a cooperative navigation system, providing theoretical guidance for the design of subsequent algorithms.

After an underwater acoustic measurement is performed by an AUV, the positioning error of the AUV in the direction of the underwater acoustic measurement is ε, and the positioning error in the direction of the underwater acoustic measurement is $\bar{\varepsilon}$; then, the positioning error of the AUV can be expressed as an ellipse error, that is, $\varepsilon = \sigma$, which is determined by the measurement accuracy of underwater acoustic measurement equipment (e.g., USBL). Assume ε_k and $\bar{\varepsilon}_k$ are the errors of a slave AUV at a time k; by taking the AUV position as the origin, the polar equation of the error ellipse can be defined as follows:

$$r^2 = \frac{\bar{\varepsilon}_k^2 \varepsilon_k^2}{\bar{\varepsilon}_k^2 \sin^2 \beta + \varepsilon_k^2 \cos^2 \beta}, \tag{8}$$

where $|r|$ is the modulus length of the error vector from the origin to any point on the error ellipse, and β is the angle between this error vector and the horizontal axis of the error ellipse.

After an interval Δt, at moment $k+1$, a slave AUV conducts another acoustic measurement. Since the observation data are one-dimensional distance data, the error can only be reduced on the vertical axis of the observation direction. As shown in Figure 2, based on the polar coordinate Equation (8) of the error ellipse, the error propagation equation for multi-AUV cooperative navigation is given by:

$$\begin{cases} \bar{\varepsilon}_{k+1}^2 = \frac{\bar{\varepsilon}_k^2 \varepsilon_k^2}{\bar{\varepsilon}_k^2 \sin^2 \gamma_{k+1} + \varepsilon_k^2 \cos^2 \gamma_{k+1}} + \zeta \cdot \Delta t \\ \varepsilon_{k+1}^2 = \varepsilon_0^2 \end{cases} \tag{9}$$

where ζ is the error propagation growth factor, and it is related to the speed measurement accuracy of a slave AUV's speed measurement equipment (e.g., DVL); $\gamma_{k+1} = |\theta_k - \theta_{k+1}|$ is the absolute difference in the azimuth angle between two adjacent acoustic measurements; ε_0 is related to the distance measurement accuracy of acoustic measurement equipment.

According to Equation (9), since the observation data only include one-dimensional distance observation data, the error in the vertical direction of the acoustic measurement continuously accumulates, so $\bar{\varepsilon}_k > \varepsilon_k$. To analyze the error propagation characteristics of the multi-AUV cooperative navigation further, this study uses different ε_k and $\bar{\varepsilon}_k$ to analyze the relationship between $\bar{\varepsilon}_{k+1}$ and γ_{k+1}, as shown in Figure 3.

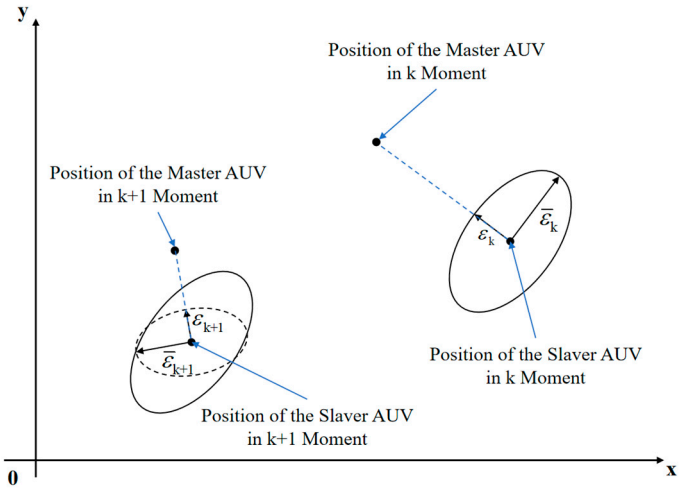

Figure 2. The multi-AUV cooperative positioning error model.

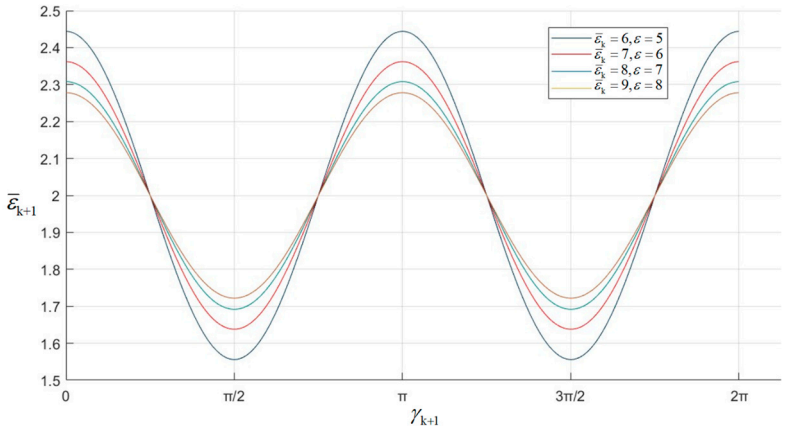

Figure 3. Propagation curves of the multi-AUV cooperative positioning error.

In Figure 3, it can be seen that when γ_{k+1} is 90 degrees or 270 degrees, the positioning error is minimized. The error analysis results are consistent with the aforementioned observability analysis.

3. Proposed Method

3.1. Cooperative Navigation Method under the Markov Decision Framework

The problems solved by reinforcement learning methods are all modeled based on the Markov decision process (MDP) [26]. The MDP represents an optimal decision-making process for dynamic stochastic systems modeled based on the Markov decision theory. The MDP is expressed as $M = (S, A, P^a_{ss'}, R^a_s)$, where$S$ represents the state space, A represents the action space, $P^a_{ss'}$ is the state transition matrix, and R^a_s is the reward function. The state set, action set, and reward function for this research problem considered in this work are defined as follows:

(1) State Set

The state set should be selected so that it fully describes the system state and is as concise as possible. A redundant state set can result in both a large number of states to be learned, consuming a significant number of computational resources, and cause the training process to not converge. As can be inferred from Equation (9), in a multi-AUV cooperative navigation system, an AUV's positioning error is mainly influenced by changes in the relative distance measurement angle. Therefore, the state set can be defined as follows:

$$S = \left\{ \hat{\theta}_k^i, \hat{d}_k^i \right\}, \tag{10}$$

where $\hat{\theta}_k^i$ is the relative bearing angle value between the master and slave AUVs, and $\hat{d}_k^i$ is the relative distance measurement value between the master and slave AUVs.

Moreover, to address the problem of a limited dimension, the state values in the state set need to be discretized;

(2) Action Set

For a multi-AUV cooperative navigation system, the action set can be defined as a subset of the set obtained after discretizing the heading angle velocity ω_k^m of a master AUV, which can be expressed by:

$$A \in \{\omega_{min}, \ldots, \omega_{max}\}, \tag{11}$$

where ω_{min} and ω_{max} are the minimum and maximum heading angle velocity values the AUV can achieve, respectively;

(3) Reward Function

The main purpose of introducing a cooperative navigation reward function is to reduce the positioning error of a slave AUV. Therefore, the theoretical positioning error of the ith slave AUV at a time k, computed by Equation (9), is considered as a cost C_k^i for an action a taken by a master AUV:

$$C_k^i = \left(\bar{\varepsilon}_k^i\right)^2 + \left(\varepsilon_k^i\right)^2. \tag{12}$$

However, it is necessary to ensure an appropriate distance between the master and slave AUVs. This distance should be neither too close, falling below the minimum safety distance between the two AUVs, nor too far, exceeding the maximum operational range of underwater communication equipment. Since the trajectory of a slave AUV is pre-planned, there is no need to consider the distance between slave AUVs. To ensure that a master AUV always maintains an appropriate distance from a slave AUV during navigation, when the distance between them is too close or far, a penalty P_k^i is imposed on the master AUV:

$$P_k^i = \begin{cases} e^{(c(d_{\min} - d_k^i) - 1)}, & d_k^i \le d_{\min} \\ 0, & d_{\min} \le d_k^i \le d_{\max}, \\ e^{(c(d_k^i - d_{\max}) - 1)}, & d_k^i \ge d_{max} \end{cases} \tag{13}$$

where c is the penalty coefficient, which is used to control penalty severity.

Combining Equations (12) and (13), the reward for an action a executed by a master AUV at a time k is obtained by:

$$R_k = -\sum_i (C_k^i + P_k^i), \quad i = 1, 2, \ldots. \tag{14}$$

3.2. Hierarchical Reinforcement Learning-Based Approach

When analyzing the collaborative navigation in a single-master single-slave AUV scenario, the master AUV trajectory planning can be undertaken based on the above-mentioned model. To reduce costs and fully leverage collaborative navigation, this study adopts a single-master multi-slave configuration. This study extends the trajectory planning method to multi-slave AUVs and incorporates the concept of hierarchical reinforcement learning, developing a trajectory planning method based on the hierarchical model and overcoming the problem of dimensionality in traditional reinforcement learning-based multi-agent decision-making.

3.2.1. Q-Learning

The reinforcement learning type used in this study is Q-learning, and the proposed algorithm represents an offline control method of temporal differences. The topology of the proposed algorithm is shown in Figure 4.

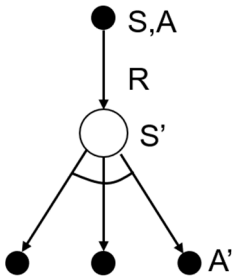

Figure 4. The topology of the Q-learning-based algorithm.

The Q-learning algorithm considers a state S' and uses the greedy method to directly select A'; namely, the maximum action a is selected to update the value function $Q(S', a)$, which is expressed by:

$$Q(S, A) \leftarrow Q(S, A) + \alpha[R + \gamma \max_{a} Q(S', a) - Q(S, A)]. \quad (15)$$

Upon completing training, the action-value function for a slave AUV, denoted by Q^*, is acquired. When a master AUV's state is initialized, optimal actions are continually selected and executed according to Equation (8) until the navigation process ends, resulting in the planned trajectory of the master AUV.

$$a^* = \underset{a' \in A(s)}{\operatorname{argmax}} Q^*(s, a'), s \in S \quad (16)$$

3.2.2. Abstract Actions

In traditional reinforcement learning, actions are perceived as instantaneous, only spanning a single timestep, and these actions are labeled as primitive actions. However, in a hierarchical model, actions can span multiple timesteps, and such macro-level actions are called abstract actions. Abstract actions denote policies that combine sequences of low-level actions into a singular, high-level action. The aim is to simplify the decision-making process for agents, reduce the number of actions they must consider, and enable policy optimization at a more abstract level.

3.2.3. Semi-Markov Decision Process

Introducing abstract actions leads to the formation of a model known as SMDP. The difference between the MDP and the SMDP is illustrated in Figure 5, where the top curve shows the state transitions in the MDP, where each state is separated by a uniform timestep, and the bottom curve represents the SMDP state transition, where there are abstract actions spanned by multiple timesteps.

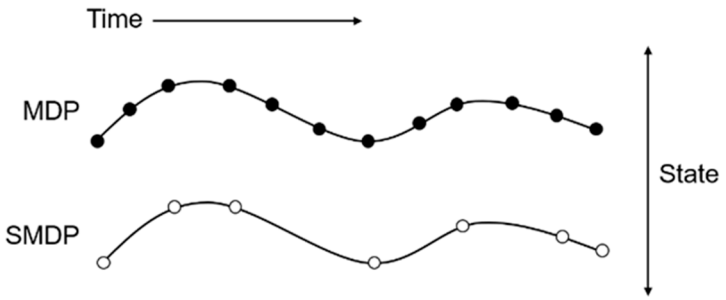

Figure 5. Illustration of the MDP and SMDP.

The SMDP can be defined by a tuple $\{S, A, P, R\}$. The primary difference between the SMDP and the MDP is that in the SMDP, actions within a set A have a duration τ. The transition probability matrix elements in P are defined by the conditional probability as follows:

$$p(s', r|s, o, \tau) = \Pr\{S_t = s'|S_{t-1} = s, A_{t-1} = o\}. \tag{17}$$

The reward for an abstract action a is defined as an accumulated reward obtained over its duration. Typically, a discount factor γ is also incorporated as follows:

$$r(s, o, s', \tau) = \mathbb{E}[\sum_{i=0}^{\tau-1} \gamma^i R_{t+i+1}|S_{t+i} = s, A_{t+i} = o]. \tag{18}$$

3.3. Trajectory Planning Method Based on the Hierarchical Model

Following the aforementioned hierarchical reinforcement learning theories, the trajectory planning task of a master AUV in a multi-slave AUV scenario is stratified. The specific task hierarchy is presented in Figure 6.

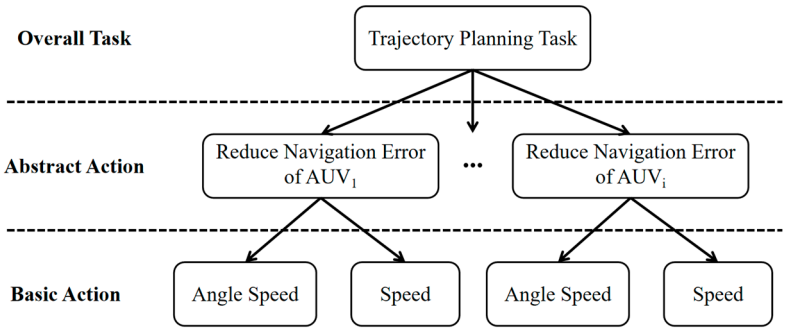

Figure 6. The collaborative navigation task's hierarchical structure.

Above, abstract actions are symbolized by o, ranging from one to n and representing the number of slave AUVs. Each abstract action's duration is set to τ, which is equivalent to the total task time. The overarching objective of a master AUV's trajectory planning is to minimize the cumulative localization errors of slave AUVs. Therefore, each slave AUV's positioning error becomes a state of the present system. Based on Equation (10), the system's state set is defined by:

$$S = \left\{\hat{\theta}_k^i, \hat{d}_k^i, \varepsilon_k^i\right\}, \quad k = 0, \tau, 2\tau, 3\tau, \ldots, \tag{19}$$

where k is an integer multiple of τ; $\hat{\theta}_k^i$ represents the relative azimuth measurement value of the ith slave AUV to the master AUV at a time k; $\hat{d}_k^i$ represents the relative distance measurement value of the ith slave AUV to the master AUV at a time k; ε_k^i is the theoretical localization error of the ith slave AUV at a time k, and it is defined by Equation (9).

During the first-tier training, a unique optimal abstract action is designed for each slave AUV within the collaborative navigation system. This action minimizes the localization error of the particular slave AUV, and its duration is fixed at τ. After incorporating abstract actions, the action set becomes:

$$A \in \{o, \omega_{min}, \ldots, \omega_{max}\}, \tag{20}$$

where $o \in \{1, 2, \ldots, n\}$ is a set of abstract actions, whose values correspond to the slave AUV numbers.

Each abstract action is accomplished through a sequence of primitive actions, specifically altering the master AUV's course angle velocity or navigation speed. A detailed analysis has revealed that each abstract action's duration represents a standard MDP, and its state space S^i is a subset of S, which is defined by:

$$S^i = \left\{\hat{\theta}_k^i, \hat{d}_k^i\right\}. \tag{21}$$

The reward for an abstract action represents a cumulative reward of its internal primitive actions. Thus, the reward generated by the change in the relative measurement angle of the ith slave AUV when the master AUV executes an abstract action o for a duration τ at a time k is calculated by:

$$R_{k+\tau}^i = -\sum_{t=0}^{\tau-1} (C_{k+t}^i + P_{k+t}^i), \quad k = 0, \tau, 2\tau, 3\tau, \ldots. \tag{22}$$

The total reward is obtained by:

$$R_{k+\tau} = \sum_{i=1}^{n} R_{k+\tau}^i = \sum_{i=1}^{n} \left\{ -\sum_{t=0}^{\tau-1} (C_{k+t}^i + P_{k+t}^i) \right\}, \quad k = 0, \tau, 2\tau, 3\tau, \ldots. \tag{23}$$

In the hierarchical model, each tier requires a decision-making strategy for the hierarchy to make appropriate choices. Since every abstract action represents a standard MDP, the decision-making policy inside an abstract action can be attained through Q-learning training. During the learning process, the master AUV conducts individual training for each slave AUV's trajectory, resulting in the corresponding action-value functions. The corresponding basic actions can be inferred from the action-value functions until the abstract action concludes, resembling the single slave AUV process, which can be expressed by:

$$a^* = \underset{a' \in A}{\operatorname{argmax}} Q^{*,i}(s, a'), s \in S. \tag{24}$$

The overarching task's decision-making strategy determines which of n obtained abstract actions should be performed. Given a limited number of slave AUVs in a single-master collaborative navigation system (typically 2–4), the general decision-making strategy is directly established using human logic and intuition. At every decision point, the master AUV should prioritize navigation for the slave AUV with the largest theoretical localization error. Accordingly, the selected abstract action should correspond to the AUV with the smallest reward as the initial condition for subsequent training:

$$o_k = \underset{i}{\operatorname{argmin}} R_k^i. \tag{25}$$

The specific steps of the proposed collaborative navigation method for a single-master multi-slave scenario are as follows:

(a) Designate the trajectories and relevant parameters for slave AUVs;
(b) For each slave AUV, use Equation (10) to determine the discretized system state set and Equation (11) to obtain the discretized action set A_i;

(c) Use the Q-learning algorithm to train the master AUV for each slave AUV; compute the instantaneous reward of the master AUV's actions using Equation (14), and eventually, obtain optimal action-value functions;
(d) Initialize the state of the master AUV, partition the sub-navigation processes, and randomly select one slave AUV along with its corresponding optimal action-value function Q^i;
(e) Use Equation (16) to select and execute an optimal action. For each slave AUV, compute the action's cost value C and calculate the total cost;
(f) Repeat Step (e) until the sub-navigation process concludes, obtaining the total cost value for each slave AUV. Thereafter, select the AUV with the highest cumulative cost value for the next sub-navigation process;
(g) Upon the completion of the final sub-navigation process, obtain the planned trajectory for the master AUV;
(h) The master AUV and multiple slave AUVs navigate according to their designated trajectories. Periodically, the master and slave AUVs acoustically communicate and measure distances between each other. The slave AUVs correct cumulative errors resulting from their navigation system outputs and perform the UKF filtering algorithm on the master AUV's position data, relative distance measurement information, and their own navigation data.

The trajectory planning method based on the hierarchical model establishes a layered model by decomposing the state space. This method eliminates non-critical states, transitioning the state space's growth from exponential to approximately linear, thus avoiding the problem of dimensionality and accelerating the solution process.

In a multi-AUV collaborative navigation system, continuous motion trajectories are segmented into fixed-interval waypoints after temporal discretization. These waypoints denote nodes where the master and slave AUVs acoustically communicate. At a time t_1, the master AUV acoustically communicates with two slave AUVs and measures their relative distances and azimuth angles. Based on these measurement data, the current system state is ascertained as shown in Figure 7. Then, the master AUV selects an optimal action a^* based on the learned optimal action-value function, transitioning the system state to S'. This process is reiterated at each subsequent acoustic communication node, finally determining the master AUV's best trajectory.

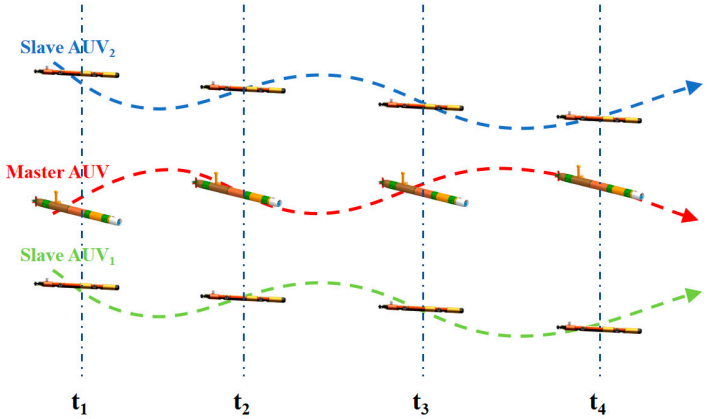

Figure 7. The multi-AUV collaborative navigation scheme.

4. Simulation Experiments

4.1. Algorithm Simulation Analysis

The simulation experiment was conducted using a cooperative navigation scenario with a single master AUV and two slave AUVs. The efficiency of the proposed trajectory planning method was analyzed for multiple slave AUVs using the changes in relative distance and the theoretical error calculated by Equation (9) as metrics. Initially, two slave

AUVs' trajectories with a uniform linear motion were defined. Slave AUV1 started from the origin (0, 0) and moved northward at a speed of 1.5 m/s, whereas slave AUV2 began from the point (150, 0) and proceeded northward at the same speed of 1.5 m/s. The master AUV embarked from the starting point (50, 50) following the trajectory planned by the proposed method at a speed of 2 m/s. The navigation duration was set to 4000 s, and acoustic measurements between the master and slave AUVs were performed every 10 s. Each abstract action lasted for 200 s, and the maximum number of training iterations was set to 1000.

In the training of the two slave AUVs with respect to the master AUV, the variations in the cost introduced by changes in the observation angle are illustrated in Figure 8.

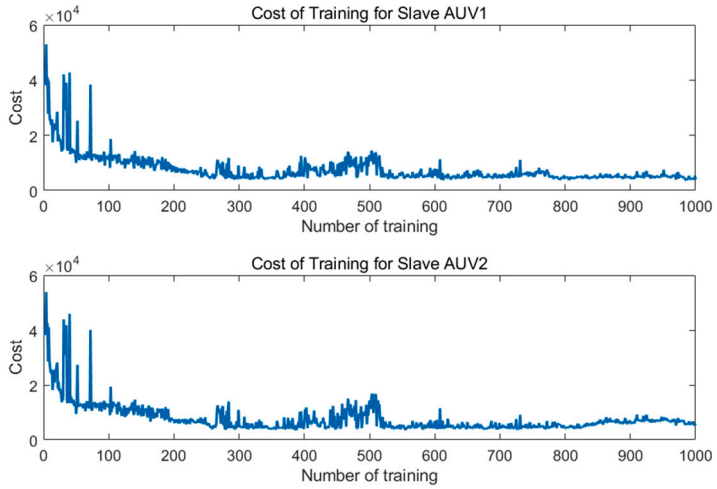

Figure 8. The changes in the cost value induced by the observation angle variations during the training process.

As shown in Figure 8, after approximately 500 training iterations, the cost value attributed to the observation angles gradually stabilized and eventually converged to its minimum value. After training, the state value function for each slave AUV was obtained. Subsequently, the action selection was performed using the decision-making approach based on hierarchical reinforcement learning. The resulting trajectory of the master AUV is depicted in Figure 9.

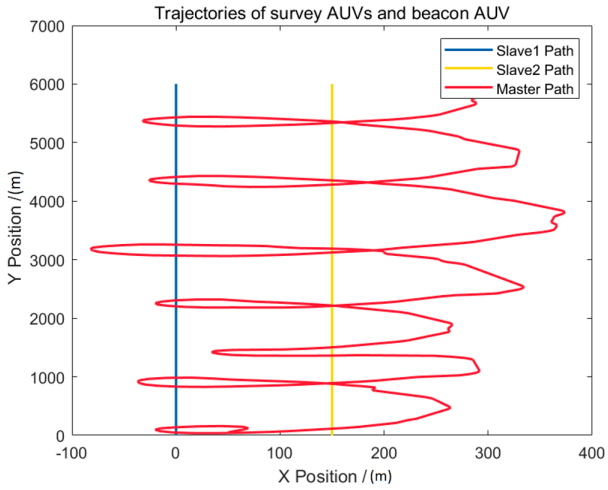

Figure 9. Trajectories of the master and slave AUVs.

In Figure 9, the red curve represents the trajectory of the master AUV, and the blue and yellow curves correspond to the trajectories of slave AUV1 and AUV2, respectively.

During the navigation period, the master AUV continuously maneuvered to minimize the localization errors of the two slave AUVs. The changes in the relative distance between the master AUV and the two slave AUVs during the entire travel are shown in Figure 10.

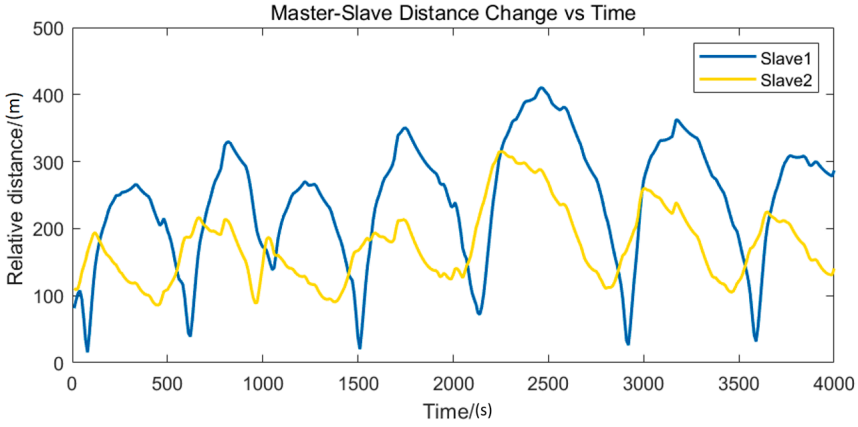

Figure 10. Changes in the relative distance between the master and slave AUVs.

As shown in Table 1, the master AUV maintained an appropriate distance from both slave AUVs during the navigation period. The changes in the theoretical localization error of the slave AUVs, calculated by Equation (9), are shown in Figure 11.

Table 1. Relative distance statistics between the master and slave AUVs.

Slave AUV	Maximum Distance (m)	Minimum Distance (m)	Average Distance (m)
AUV1	410.096	15.387	223.938
AUV2	315.016	86.083	192.137

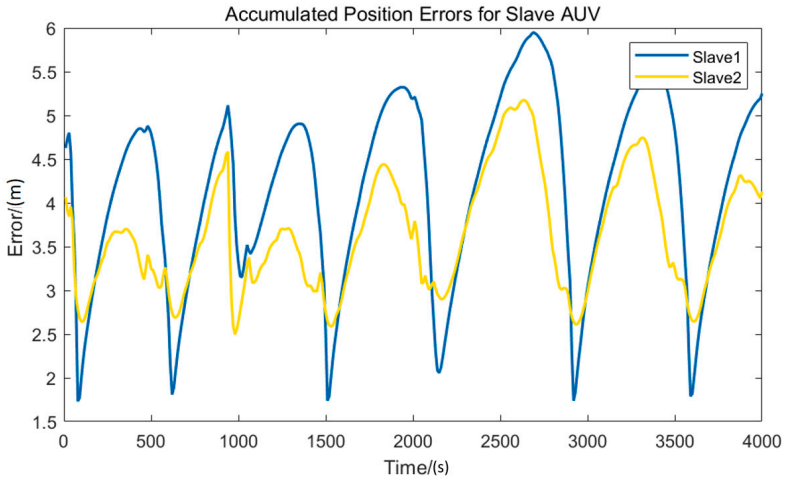

Figure 11. Changes in the theoretical localization error of the two slave AUVs.

Statistical data on the theoretical localization errors of the slave AUVs are presented in Table 2.

Table 2. Theoretical localization error statistics of the two slave AUVs.

Slave AUV	Maximum Distance (m)	Minimum Distance (m)	Average Distance (m)
AUV1	5.950	1.737	4.271
AUV2	5.174	2.497	3.836

Based on the data presented in Table 2, the theoretical localization errors of the two slave AUVs were very close during the entire navigation period. This indicated that the trajectory of the master AUV could effectively consider both slave AUVs, efficiently reducing their localization errors. In this experiment, the state set size generated by each slave AUV was 180, which resulted in a total of 360 states of the two slave AUVs. Therefore, directly solving this problem using the Q-learning-based method could result in a state set size of 32,400. However, using the trajectory planning method based on the hierarchical model could reduce the state set size by 90 times, thus reducing the time and storage space consumption.

In uniform trajectory planning, the selection of the master AUV's speed is crucial for error control. Namely, using an inappropriate speed might cause the training process to diverge. Therefore, a flexible method is needed to adjust the speed of the master AUV in real-time. A variable-speed trajectory planning method allows a master AUV to change its navigation speed within a certain range in real-time. This not only allows better error control but also avoids training process divergence due to inappropriate speed selection.

The two subsets of the action set were defined as follows:

$$\begin{cases} \mathrm{A}_1 \in \{\omega_{min}, \ldots, \omega_{max}\} \\ \mathrm{A}_2 \in \{v_{min}, \ldots, v_{max}\} \end{cases} \tag{26}$$

where A_1 is the action set after discretizing the yaw angle speed; A_2 is the action set after discretizing the navigation speed; ω_{min} and ω_{max} denote the minimum and maximum yaw angle speeds of the master AUV, respectively; v_{min} and v_{max} are the minimum and maximum navigation speeds of the master AUV, respectively.

The final action set was obtained as a Cartesian product of the two action sets as follows:

$$\mathrm{A} = \mathrm{A}_1 \times \mathrm{A}_2. \tag{27}$$

4.2. Simulation Parameter Settings

Based on the previous analysis results, the navigation equipment parameters of the master and slave AUVs were set, as shown in Table 3:

Table 3. Parameters of the master and slave AUVs.

Parameter	Master AUV	Slave AUV
Speed measurement noise (m/s)	0.5	1.5
Angle speed measurement noise (rad/s)	0.1	0.5
Acoustic measurement noise (m)	8	8
Acoustic measurement period (s)	10	10

Based on the parameters' values presented in Table 3, the navigation equipment measurement accuracy of the slave AUV was relatively poor. This was because the slave AUVs had equipment with a lower accuracy to achieve cost reduction. Then, the cooperative navigation algorithms were used to reduce their localization errors.

To use the Q-learning algorithm for trajectory planning for the master AUV, the relevant state quantities were discretized to obtain discrete state and action sets. The action quantity was obtained by Equation (10), and the state quantity discretization parameters are given in Table 4.

Table 4. The state quantity discretization parameters.

Action and State	Discrete Quantity	Number
Distance measurement azimuth angle (°)	[0, 10), [10, 20), . . ., [350, 359)	36
Relative distance (m)	[0, 100), [100, 300), [300, 600) [600, 900), [900, ∞)	5

As shown in Table 4, the relative distance measurement azimuth angle between the master and slave AUVs was discretized into thirty-six intervals of 10°, each of which considered one state. The minimum distance between the master and slave AUVs was set to 100 m, and their maximum distance, which was obtained based on the effective range of the acoustic measurement equipment, was 900 m, and there were a total of five states.

The action of the master AUV was the yaw rate. Considering that the actual AUV's maximum yaw rate was 0.08 rad/s and the maximum speed was 2.5 m/s, the yaw rate action set A_1 was selected as follows:

$$A_1 = [-0.08, -0.05, -0.03, 0.00, -0.03, -0.05, 0.08]. \quad (28)$$

Similarly, the speed action set A_2 was defined as:

$$A_2 = [1,\ 1.5,\ 2,\ 2.5]. \quad (29)$$

The reward function during the Q-learning algorithm training process was given by Equation (14), and its parameters are given in Table 5.

Table 5. The reward function parameters.

Parameter	Symbol	Value
Error propagation factor	ξ	0.1
Acoustic measurement accuracy	ε_0	1
Punish coefficient	c	0.06

By using the parameters in Table 5, the Q-learning algorithm-related parameters were determined. These parameters mainly included the learning step size, decay factor, and exploration rate, and their values are presented in Table 6.

Table 6. The Q-learning algorithm parameters.

Parameter	Symbol	Value
Study step	α	0.015
Decay factor	γ	0.9
Exploration rate	ϵ	0.1

Next, the UKF filter parameters were set, as shown in Table 7.

Table 7. The navigation simulation experimental parameters.

	Speed Measurement Noise (m/s)	Angle Speed Measurement Noise (Rad/s)	Acoustic Measurement Noise (m)
Master AUV	$N(0, 0.5^2)$	$N(0, 0.1^2)$	$N(0, 8^2)$
Slave AUV	$N(0, 1.5^2)$	$N(0, 0.5^2)$	

As shown in Table 7, the selected master AUV-related measurement noise was small; namely, based on the actual situation of a cooperative navigation system, the master AUV

was equipped with high-precision, high-cost navigation equipment. In contrast, the slave AUVs had low-precision, low-cost navigation equipment, so their measurement noise was large. Finally, the master and slave AUVs used the same acoustic measurement equipment, so their acoustic measurement noises were the same. All of the above-presented noises were zero-mean Gaussian white noise.

4.3. Trajectory Planning Analysis

After setting the simulation parameters, a simulation experiment was performed in the cooperative navigation scenario with one master AUV and three slave AUVs. The trajectories of the three slave AUVs were all serpentine search curves. Based on the experimental results, the performance of the proposed method for the cooperative navigation system with multiple slave AUVs was analyzed using the curve routes.

First, the trajectory planning process was conducted. The three slave AUVs started from the points $(-250, 0)$, $(0, 0)$, and $(250, 0)$ and performed a uniform linear motion with a navigation speed of 1.5 m/s. The simulation time was 4000 s. The master AUV used the hierarchical reinforcement learning trajectory planning method with a navigation speed of 2.5 m/s, and the maximum number of training epochs was set to 1000.

After 1000 training epochs, the cost changes caused by the observation angle changes were calculated, as shown in Figure 12.

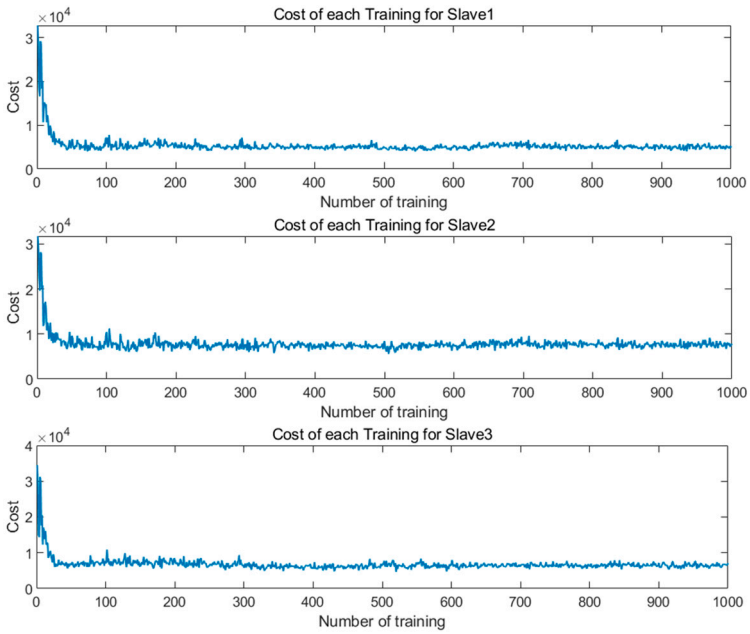

Figure 12. The cost changes with the observation angle value.

As shown in Figure 12, after about 30 training epochs, the cost caused by the observation angle changes gradually converged to the minimum value. During the training process, the master AUV continuously explored new decisions, so the cost value fluctuated. After the training was completed, the action-value function table was obtained, and then the master AUV trajectory was planned, as shown in Figure 13.

For the trajectories presented in Figure 13, the relative distance changes between the master and slave AUVs are shown in Figure 14.

As displayed in Figure 14, the maximum distance between the master AUV and slave AUV1 was 406.35 m, with an average value of 191.33 m; the maximum distance between the master AUV and slave AUV2 was 625.29 m, with an average value of 305.31 m; and the maximum distance between the master AUV and slave AUV3 was 596.52 m, with an average value of 297.65 m. During travel, the distance between the master and slave AUVs

always remained within the maximum operational distance of the acoustic communication equipment of 900 m.

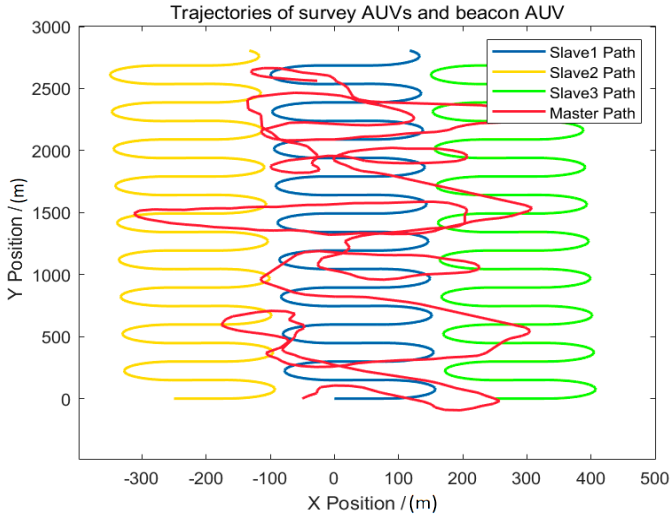

Figure 13. The master and slave AUVs' trajectories.

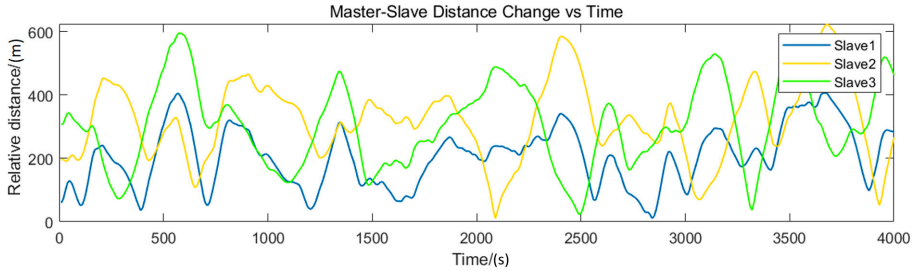

Figure 14. Relative distance changes between the master and slave AUVs.

The theoretical localization error changes of the slave AUVs during travel are shown in Figure 15.

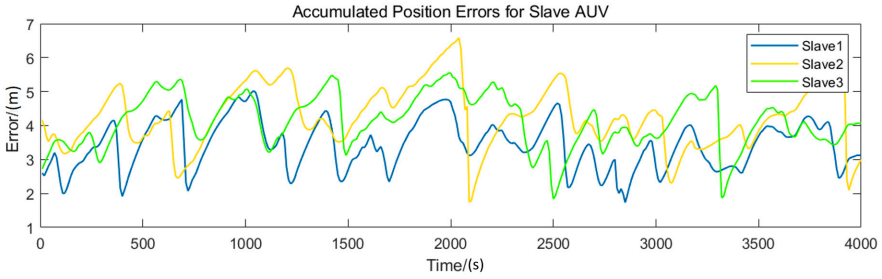

Figure 15. The theoretical localization errors of the slave AUVs.

As displayed in Figure 15, the minimum theoretical error of the three slave AUVs at the start of the navigation process was 2.84 m. As the navigation process progressed, the theoretical localization errors of the three slave AUVs continuously increased, finally stabilizing at approximately 3.5 m. This indicated that the trajectory of the master AUV could meet the observable conditions and keep the theoretical localization error of the slave AUVs bounded.

4.4. UKF Filtering Analysis

The trajectories of the slave AUVs' direct dead reckoning are presented in Figure 16.

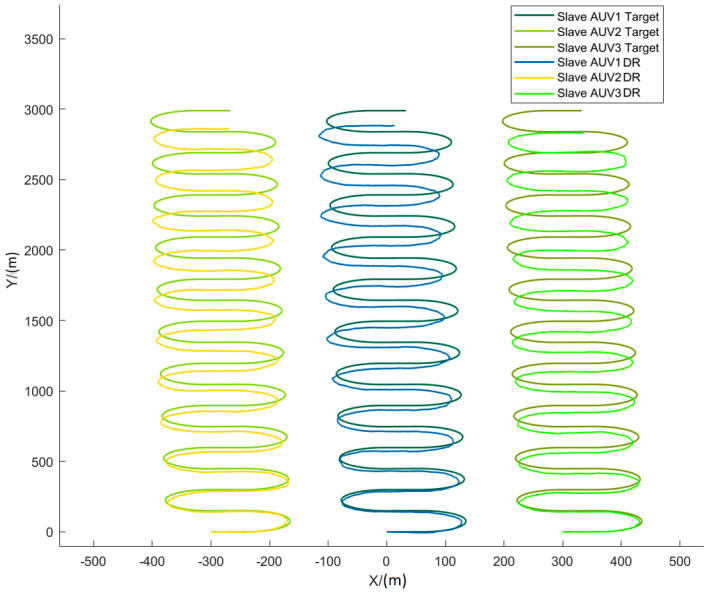

Figure 16. The three slave AUVs' dead reckoning trajectories.

As illustrated in Figure 16, as the navigation process progressed, the dead reckoning trajectories of the slave AUVs gradually deviated from their actual curves. Among the three slave AUVs, the trajectory deviation of slave AUV1 was the most significant. Furthermore, the dead reckoning errors of the three slave AUVs mainly manifested in the *Y*-axis direction. To analyze the proposed method further, this study conducted 100 navigation simulation experiments, and the positioning error statistical data are shown in Table 8.

Table 8. The DR navigation test error statistical results.

Slave AUV	Average RMS Error		Average Relative Error (m)
	X-Axis Distance (m)	*Y*-Axis Distance (m)	
AUV1	22.20	184.46	322.04
AUV2	21.47	183.44	315.18
AUV3	22.27	179.80	314.63

Based on the results in Table 8, the positioning error mainly originated from the error in the *Y*-axis direction. This was because the displacement of the slave AUVs in the *Y*-axis direction was longer. However, after using the UKF filtering algorithm for cooperative navigation, after one navigation calculation, the trajectories of the master and slave AUVs were recalculated, as shown in Figure 17.

The positioning errors of the slave AUVs during the navigation period in the navigation test are shown in Figure 18.

As shown in Figure 18, in this navigation experiment, when only relying on dead reckoning, the positioning errors of the slave AUVs continuously grew and diverged after the start of navigation. At the end of the navigation process, the errors reached their maximum values. The maximum positioning errors of the three slave AUVs were 1058.16 m, 1058.46 m, and 1076.43 m. After using the UKF algorithm in the navigation calculation, the errors of the three slave AUVs could stabilize within a certain range. The positioning errors showed a significant fluctuation at a time of approximately 2800 s during navigation. During the navigation period, the maximum positioning errors of the three slave AUVs were 361.92 m, 394.55 m, and 360.73 m, respectively, with average values of 252.42 m, 240.16 m, and 239.24 m, respectively.

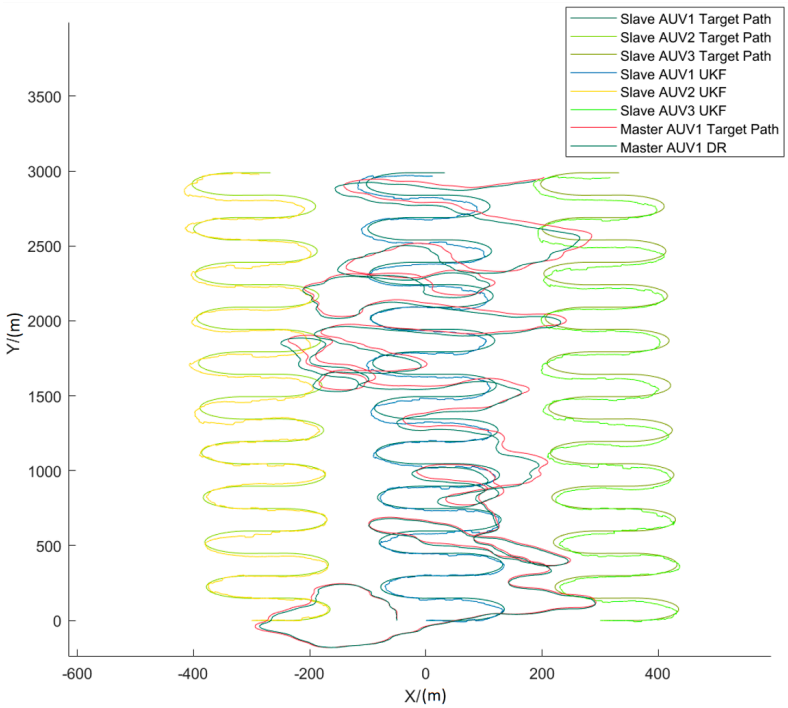

Figure 17. The AUVs' trajectories after UKF filtering.

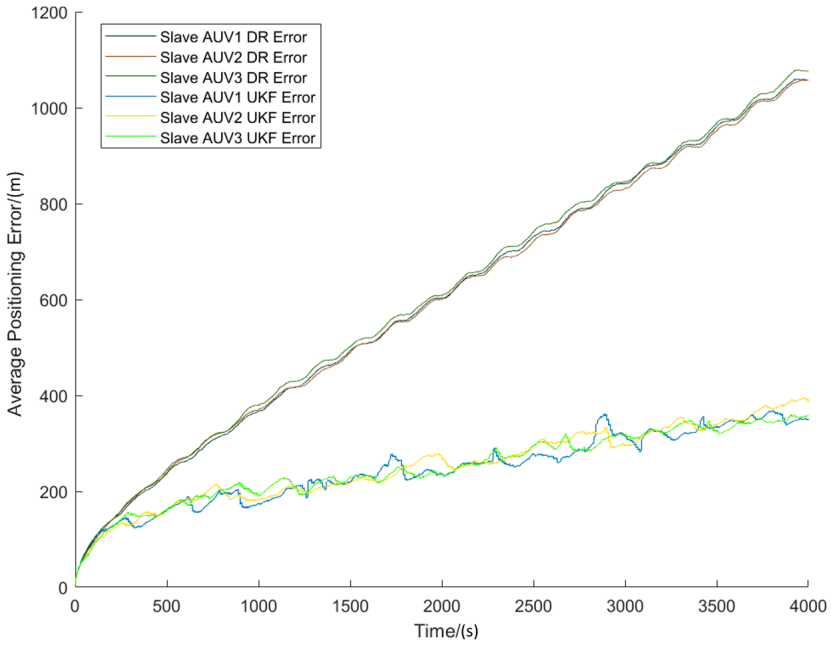

Figure 18. The slave AUVs' positioning errors after the UKF.

Next, statistical data were obtained after 100 navigation simulation experiments to analyze the proposed method further, and the obtained results are shown in Table 9.

Based on the results in Table 9, the positioning errors of the three slave AUVs mainly originated from the error in the *Y*-axis direction, this can be considered as the result of the combination of the cumulative error and the aforementioned elliptic error. Compared to the results obtained when solely relying on dead reckoning, after using the proposed method, the positioning errors were reduced by approximately four times. This indicated that applying the proposed method could significantly reduce the positioning error of the slave AUVs, thereby enhancing the positioning performance of the multi-AUV cooperative navigation system.

Table 9. The UKF navigation error statistics.

Slave AUV	Average RMS Error		Average Relative Error (m)
	X-Axis Distance (m)	*Y*-Axis Distance (m)	
AUV1	15.75	46.35	80.37
AUV2	16.45	51.60	94.32
AUV3	14.59	48.64	80.51

5. Conclusions

Considering a master-slave multi-AUV cooperative navigation system, this paper proposes a multi-AUV cooperative navigation method based on hierarchical reinforcement learning. The proposed method adopts a single-master multi-slave structure. The trajectories of the slave AUVs are pre-planned according to the navigation task requirements. The algorithm then plans the trajectory for the master AUV, reducing the observation and positioning errors of the slave AUVs. The proposed method divides the entire cooperative navigation process into two parts, the trajectory planning process and the navigation calculation process. In the trajectory planning process, the MDP model of the cooperative navigation problem is constructed, and the abstract actions of the single-master multi-slave AUV cooperative navigation system are defined. Based on the hierarchical reinforcement learning-based algorithm, a trajectory planning method for the master AUV is designed. In the navigation calculation process, the trajectory planning results are combined with the UKF filtering method to realize a complete cooperative navigation method process. The proposed method is verified by navigation simulation experiments using two different cooperative navigation scenarios. By introducing a hierarchical structure and abstract actions, the proposed method achieves task decomposition and hierarchical management during the trajectory planning phase of multiple slave AUVs. Using the hierarchical structure effectively reduces the planning complexity of the master AUV, and abstract actions decrease the size of the state space, making the cooperative navigation system more efficient and flexible. The proposed hierarchical approach better addresses the complexity and large-scale problem of multi-AUV cooperative navigation tasks, significantly improving the performance and task completion quality of the multi-AUV cooperative navigation system, which is proposed for the first time in related researches, and has excellent performance in multi-AUV cooperative navigation problems. In addition, the proposed method also has a certain degree of adaptability to outliers, making the multi-AUV cooperative navigation system more robust and stable in complex underwater environments.

Author Contributions: Conceptualization, Z.Z. and L.Z.; methodology, Z.Z., D.W. and R.R.; software, D.W.; validation, L.Z. and L.L.; formal analysis, Z.Z. and W.G.; investigation, S.B.; resources, L.Z.; data curation, S.B.; writing—original draft preparation, Z.Z.; writing—review and editing, Z.Z., L.Z. and L.L.; visualization, D.W. and W.G.; supervision, L.Z.; project administration, L.Z. and L.L.; funding acquisition, L.Z. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the National Natural Science Foundation of China (No. 52371339 and 52001259), Shenzhen Science and Technology Program under Grant JCYJ20210324122010027 and JCYJ20210324122406019, Local Science and Technology Special fundation under the Guidance of the Central Government of Shenzhen under Grant 2021Szvup111, Science and Technology on Avionics Integration Laboratory and Aeronautical Science Foundation of China under Grant 201955053003, the China Postdoctoral Science Foundation under Grant 2020M673484, National Research and Development Project under Grant 2021YFC2803000, the National Key Research and Development Program of China (Grant No. 2020YFB1313200, 2020YFB1313202, 2020YFB1313204).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, J.; Si, Y.; Chen, Y. A Review of Subsea AUV Technology. *J. Mar. Sci. Eng.* **2023**, *11*, 1119. [CrossRef]
2. Lambert, W.; Miller, L.; Brizzolara, S.; Woolsey, C. A Free Surface Corrected Lumped Parameter Model for Near-Surface Horizontal Maneuvers of Underwater Vehicles in Waves. *Ocean. Eng.* **2023**, *278*, 114364. [CrossRef]
3. Mendes, P.; Batista, P.; Oliveira, P.; Silvestre, C. Cooperative Decentralized Navigation Algorithms Based on Bearing Measurements for Arbitrary Measurement Topologsies. *Ocean. Eng.* **2023**, *270*, 113564. [CrossRef]
4. Zhao, Y.; Xing, W.; Yuan, H.; Shi, P. A Collaborative Control Framework with Multi-Leaders for AUVs Based on Unscented Particle Filter. *J. Frankl. Inst.* **2016**, *353*, 657–669. [CrossRef]
5. Edwards, D.B.; Bean, T.A.; Odell, D.L.; Anderson, M.J. A Leader-Follower Algorithm for Multiple AUV Formations. In Proceedings of the 2004 IEEE/OES Autonomous Underwater Vehicles, Sebasco, ME, USA, 17–18 June 2004; pp. 40–46.
6. Forsgren, B.; Vasudevan, R.; Kaess, M.; McLain, T.W.; Mangelson, J.G. Group-*k* Consistent Measurement Set Maximization for Robust Outlier Detection. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; pp. 4849–4856.
7. Guo, Y.; Xu, B.; Wang, L. A Robust SINS/USBL Integrated Navigation Algorithm Based on Earth Frame and Right Group Error Definition. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 8504716. [CrossRef]
8. Lee, K.; Johnson, E.N. Robust Outlier-Adaptive Filtering for Vision-Aided Inertial Navigation. *Sensors* **2020**, *20*, 2036. [CrossRef] [PubMed]
9. Lu, J.; Chen, X.; Luo, M.; Zhou, Y. Cooperative Localization for Multiple AUVs Based on the Rough Estimation of the Measurements. *Appl. Soft Comput.* **2020**, *91*, 106197. [CrossRef]
10. Wang, W.; Xu, Y. A Modified Residual-Based RAIM Algorithm for Multiple Outliers Based on A Robust MM Estimation. *Sensors* **2020**, *20*, 5407. [CrossRef] [PubMed]
11. Bai, M.; Huang, Y.; Chen, B.; Yang, L.; Zhang, Y. A Novel Mixture Distributions-Based Robust Kalman Filter for Cooperative Localization. *IEEE Sens. J.* **2020**, *20*, 14994–15006. [CrossRef]
12. Li, Q.; Ben, Y.; Naqvi, S.M.; Neasham, J.A.; Chambers, J.A. Robust Student's T-Based Cooperative Navigation for Autonomous Underwater Vehicles. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1762–1777. [CrossRef]
13. Bo, X.; Razzaqi, A.A.; Yalong, L. Cooperative Localisation of AUVs based on Huber-Based Robust Algorithm and Adaptive Noise Estimation. *J. Navigation* **2019**, *72*, 875–893. [CrossRef]
14. Sun, C.; Zhang, Y.; Wang, G.; Gao, W. A Maximum Correntropy Divided Difference Filter for Cooperative Localization. *IEEE Access* **2018**, *6*, 41720–41727. [CrossRef]
15. Zhang, L.; Qu, J.; Pan, G.; Wang, Y. Analyzing of Cooperative Locating Error and Formation Configuration of AUV Based on Geometric Interpretation. *J. Northwestern Polytech. Univ.* **2020**, *38*, 755–765. [CrossRef]
16. Chiarella, D. Towards Multi-AUV Collaboration and Coordination: A Gesture-Based Multi-AUV Hierarchical Language and A Language Framework Comparison System. *J. Mar. Sci. Eng.* **2023**, *11*, 1208. [CrossRef]
17. Majid, M.H.A.; Yahya, M.F.; Siang, S.Y.; Arshad, M.R. Cooperative Positioning of Multiple AUVs for Underwater Docking: A Framework. In Proceedings of the Colloquium on Robotics, Unmanned Systems and Cybernetics, Pekan, Malaysia, 20 November 2014; p. 1.
18. Zhang, L.; Li, Y.; Liu, L.; Tao, X. Cooperative Navigation Based on Cross Entropy: Dual Leaders. *IEEE Access* **2019**, *7*, 151378–151388. [CrossRef]
19. Li, Q.; Naqvi, S.M.; Neasham, J.; Chambers, J. Robust Cooperative Navigation for AUVs Using the Student's t Distribution. In Proceedings of the IEEE 2017 Sensor Signal Processing for Defence Conference (SSPD), London, UK, 6–7 December 2017; pp. 1–5.
20. Zheng, K.; Jiang, Y.; Li, Y. Passive Localization for Multi-AUVs by Using Acoustic Signals. In Proceedings of the 14th International Conference on Underwater Networks & Systems, Atlanta, GA, USA, 23–25 October 2019; pp. 1–5.
21. Yoshihara, T.; Ebihara, T.; Mizutani, K.; Sato, Y. Underwater Acoustic Positioning in Multipath Environment Using Time-of-flight Signal Group and Database Matching. *Jpn. J. Appl. Phys.* **2022**, *61*, SG1075. [CrossRef]
22. Franchi, M.; Bucci, A.; Zacchini, L.; Ridolfi, A.; Bresciani, M.; Peralta, G.; Costanzi, R. Maximum a posteriori estimation for AUV localization with USBL measurements. *IFAC-PapersOnLine* **2021**, *54*, 307–313. [CrossRef]
23. Li, J.H.; Lee, P.M. A Neural Network Adaptive Controller Design for Free-pitch-angle Diving Behavior of An Autonomous Underwater Vehicle. *Robot. Auton. Syst.* **2005**, *52*, 132–147. [CrossRef]
24. Zhang, T.; Chen, L.; Li, Y. AUV Underwater Positioning Algorithm Based on Interactive Assistance of SINS and LBL. *Sensors* **2015**, *16*, 42. [CrossRef] [PubMed]
25. Ren, R.; Zhang, L.; Liu, L.; Wu, D.; Pan, G.; Huang, Q.; Zhu, Y.; Liu, Y.; Zhu, Z. Multi-AUV Cooperative Navigation Algorithm Based on Temporal Difference Method. *J. Mar. Sci. Eng.* **2022**, *10*, 955. [CrossRef]
26. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2014.