Journal of
*Marine Science and Engineering*

MDPI

*Article*

# Towards Multi-AUV Collaboration and Coordination: A Gesture-Based Multi-AUV Hierarchical Language and a Language Framework Comparison System

**Davide Chiarella**

Institute of Computational Linguistics—National Research Council, Via E. De Marini 6, 16149 Genova, Italy; davide.chiarella@cnr.it; Tel.: +39-010-6475205

**Abstract:** The underwater environment is a harmful environment, yet one of the richest and least exploited. For these reasons the idea of a robotic companion with the task of supporting and monitoring divers during their activities and operations has been proposed. However, the idea of a platoon of robots at the diver's disposal has never been fully addressed in these proposals due to the high cost of implementation and the usability, weight and bulk of the robots. Nevertheless, recent advancements in swarm robotics, materials engineering, deep learning, and the decreasing cost of autonomous underwater vehicles (AUVs), have rendered this concept increasingly viable. Therefore, this paper introduces, in the first part, a novel framework that integrates a revised version of a gesture-based language for underwater human–robot interaction (Caddian) based on insights gained from extensive field trials. The newly introduced objective of this framework is to enable the cooperation and coordination of an AUV team by one or more human operators, while allowing a human operator to delegate a robot leader to instruct the other robotic team members. The work, in the second part, provides an evaluation of the new language proposed thanks to a fifty million sentence corpus and describes a comparison framework, which is used to estimate it with respect to other existing underwater human–robot interaction languages.

**Keywords:** gesture-based language; underwater human–robot interaction; multi-AUV collaboration; language corpora and resources



**Citation:** 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. https://doi.org/10.3390/jmse11061208

Academic Editor: Rafael Morales

Received: 20 April 2023
Revised: 31 May 2023
Accepted: 6 June 2023
Published: 10 June 2023



## 1. Introduction

The underwater domain is a hazardous environment [1], yet it is one of the richest and least exploited. Dive operators use life-support equipment and have limited navigational and communication capabilities with each other and with the surface while they are constantly exposed to the harsh environment during their activities. To cope with these difficulties, divers usually work at least in pairs with so-called buddy diving [2,3]. For these reasons, over the last twenty years, the idea of a robotic companion to divers has been explored and various solutions have been proposed. In this regard, the underwater environment also poses unique challenges when it comes to developing autonomous robots that can collaborate with human operators. In this scenario, one of the most challenging problems is the communication between human and robot and robot to robot. In fact, the methods of communication used for interaction on the surface cannot be applied or used in the underwater environment—radio/wireless technologies are not so efficient due to the attenuation of radio waves, and infrared solutions suffer from the same drawback. This is the reason why acoustic modems are usually employed, but they are power-intensive, have a low bandwidth and have a long transmission delay [4]. In addition, acoustic signaling for environmental and remediation tasks can be perceived as invasive by marine species and have deleterious effects on their well-being. On a different note, keyboards, touchscreens and tablets also do not function properly underwater and require additional marinization, and, if not an on-board solution, also need a tether or close proximity. Having

said that, research has turned towards the use of computer vision to implement underwater communication between humans and robots.

In recent years, many human–robot underwater communication approaches have been proposed that make use of computer vision. The most widely used is undoubtedly the use of gestures, for two main reasons: the first consists of the fact that for humans, the use of gestures during communication comes naturally; the second stems directly from the first since all over the world the diving community normally uses gestures to communicate underwater [5–8]. This paper refers particularly to earlier papers initially presented in [9] and later expanded in [10].

This work extends the results presented in the aforementioned articles. To the best of our knowledge, these are the only studies that have focused on gesture-based language, defined as a formal language [11], and the underwater human–robot interaction language (UHRI) herein described, Caddian. The latter has proven its usability and has already been tested in several field missions, described in [12] and in [13], and on three different robotic underwater vehicles, namely BUDDY AUV [14–16], R2 ROV and e-URoPe [17] (see Figure 1). Moreover, the language provides a public dataset [12,18] containing images of divers' gestures and poses, allowing the scientific community to work on optimising the framework [19–21].

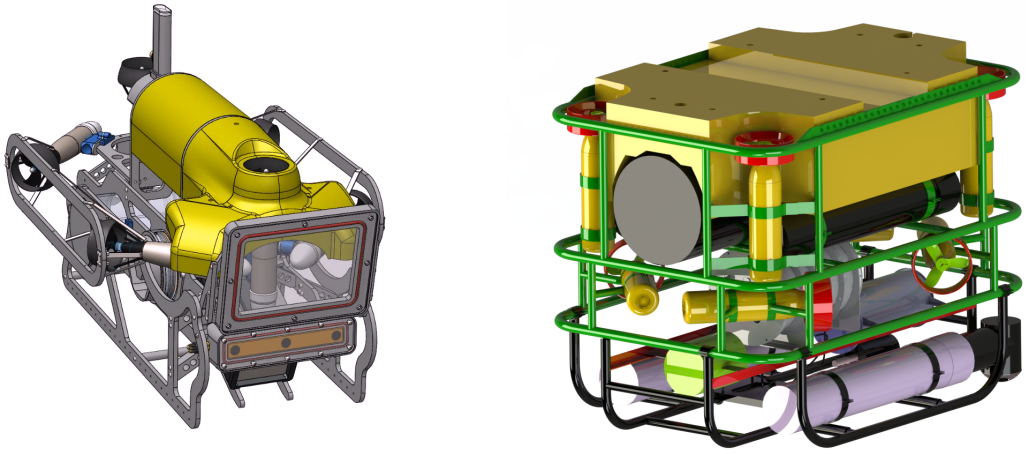

**Figure 1.** Two of the main underwater autonomous vehicles (AUVs) used with the Caddian language to assist with diver activities: on the left BUDDY AUV by UNIZEG-FER and on the right R2 by CNR.

In the existing body of literature, the development of a language that promotes effective communication and collaboration between humans and robots has been recognised as a significant challenge in the field of human–robot interaction (HRI). In this paper, we contend that the proposed extension to Caddian offers notable advantages for robotics research. The first advantage lies in its ability to facilitate effective communication between humans and robots. Traditional programming languages used for robot control often exhibit complexity, making them challenging for non-experts to comprehend. In contrast, Caddian has been specifically designed to be intuitive and user-friendly, even for individuals without technical expertise. This design approach ensures that robots can be controlled more effortlessly and efficiently, even by those without a background in robotics. The second advantage is that it can enable better coordination between humans and robots, addressing one of the key challenges in HRI, which is to ensure that robots and humans can work together safely and efficiently. Moreover, an HRI language that enables robots to communicate with humans in a clear and understandable way can help to reduce the risk of accidents and misunderstandings. This can be especially important in applications such as assisted repair and maintenance tasks underwater, where robots and humans often have to work in close proximity.

The third advantage is that it can enable robots to work together more effectively: in many applications, such as search and rescue missions, it is often necessary or desired for

multiple robots to work together to accomplish a task. In this regard, the extension to the language which we propose enables robots to coordinate their actions and communicate with each other, helping to improve their performance and increase their efficiency. This can be especially important in applications where time is critical, such as emergency response scenarios.

In summary, the addition of a section on collaboration and coordination in the Caddian UHRI language brings substantial benefits to the field of robotics. It facilitates efficient communication between humans and robots, improves coordination between humans and robots, and enhances the ability of robots to work together and adapt to dynamic situations.

In particular, this paper expands on the existing language and makes the following contributions:

- It updates and revises the Caddian language into the Caddian core language where:
  - all revisions and expansions are based on feedback from the diving community and results from trials;
  - revisions simplify some part of the language;
  - expansions add robot-to-human communication capabilities to the language;
- It involves the design of a new multi-AUV framework for cooperation and coordination of an AUV team and adds relevant components to the language;
- It proposes a hierarchical schema for designating a robot leader among AUVs;
- The proposed UHRI framework uses gestures as a means of communication, but it is agnostic, so the back-end can be used with other front-ends that recognise other means of communication besides gestures;
- It provides a Caddian corpus consisting of fifty million sentences in Caddian and evaluates it to enable researchers to know the main features of the language model and optimise the gesture recognition front-end accordingly;
- It proposes an evaluation framework for UHRI systems (evaluation criteria for UHRI-ECU);
- It compares, using ECU, the expanded language with other existing underwater human–robot interaction languages, thus providing an overview of the current state of UHRI languages.

## 2. Related Work

In order to implement underwater communication between humans and robots using computer vision, two aspects must be considered: the first is diver detection, tracking and understanding of body position; the second is the creation of an effective and robust language that can be memorised, and, at the same time, has a high expressive power along with a set of messages that can be used nimbly during emergency situations. These two concepts were well illustrated in [22] and then in [23]. Underwater communication can, in fact, be divided into a front-end problem to identify and recognise divers' gestures or more generally the communication medium, and a back-end problem that has to do with the interpretation of language. In this perspective, back-ends have the advantage that they can be combined with different front-ends and are agnostic towards the way symbols are perceived and recognised by the front-end.

Regarding the first point, front-ends to detect and recognise diver gestures have a quite large literature. In 2007, Sattar and Dudek [24] proposed an algorithm for tracking divers using periodic motion to recognise undulating flipper motion connected to typical gaits. In 2011, Buelow and Birk [25] developed an approach for gesture detection with a monocular camera on a moving vehicle. In 2013, the identification and tracking of a diver using a high-frequency forward-looking sonar was considered by De Marco et al. [26]. In 2015, Chavez et al. [27] described a visual technique for diver detection in the context of human–robot interaction which outperformed conventional generative tracking techniques, using a modified nearest class mean forests algorithm where a discriminative model was trained by computing picture attributes of the target (diver) and underwater environment. In 2017, Islam and Sattar [28], using both spatial-domain and frequency-domain data on human

swimming movements, presented an algorithm allowing an autonomous underwater robot to visually detect and follow its partner human diver. In addition, in 2017, a broad framework for detecting and estimating the diver's body posture based on point clouds produced by a stereo camera was provided by Chavez et al. [29]. In 2019, Xia and Sattar [30] developed a method for autonomous underwater robots to visually recognise and identify divers, allowing an underwater robot to identify a human team leader and detect several divers in a visual context. In 2021, Remmas et al. [31] presented a study using data-fusion of various data from inexpensive sensors to improve perception in dynamic and unstructured ocean environments of divers. In 2021, Jiang et al. [32] also proposed a gesture tracking method for underwater human–robot interaction based on fuzzy control.

However, there is not as much literature on UHRI languages, the associated communication protocols and back-ends for their interpretation [23]. The first forerunner was Robochat [33], developed in 2007 and defined via a Backus–Naur form (BNF) [34]. It uses fiducial markers to communicate with the AUV and supports macro definition and execution, the if-else statement, and the indexed iterator statement. While RoboChat offered a novel solution to the issue of underwater communication, the language, according to one of its authors [35], suffered from two critical weaknesses. First, many marker cards had to be safely carried during the dive because each language token required its own marker. This made it difficult for the diver to recall all of the tags, and also made the equipment bulky when used for tasks such as mapping the environment, inspecting underwater cables or monitoring wrecks. Second, it is not intuitive for a naive commercial diver to comprehend the symbol-to-instruction mapping. In fact, one of the key aims of the UHRI is to develop robot–diver interactions that do not need the diver to diverge considerably from typical ways of communication. Since hand gestures are one of the most natural ways for humans to communicate [36], as well as being the traditional mode of communication between divers, it is not surprising that, in the following years, gestures have become one of the most extensively used methods for underwater human–robot interaction (UHRI)—in fact, divers, who have always been confronted with the underwater world, use body language and a small set of gestures to express meaningful information between them [5–8].

Early work in this regard was undertaken in 2014 by Menix et al. [37] in order to interpret via a hidden Markov model (HMM) both static and dynamic diver hand signals using a real-time video feed. The main purposes were determining the most appropriate feature vector describing diver gestures and the optimal HMM parameters.

In 2015, Chiarella et al. [9] described the definition of a communication language based on consolidated and standardised diver gestures commonly employed during professional and recreational dives, called Caddian, together with the corresponding communication protocol. The language, developed for the European FP7 project CADDY [38], is described through a context-free grammar [11,34] and an initial list of gestures. Later in 2018, the work reported in [9] after completion of the project [13], was extended and completed [10]. The language, in its final version, has more than forty messages, is backward compatible with the current gestures used by divers, and it provides divers with a way to communicate complex messages (i.e., missions) to their robotic buddies.

In 2019, Islam et al. [35] proposed a human–robot communication framework based on hand gestures without the use of a grammar, in which instructions are atomically mapped to tuples of hand gestures (both the left and right hands). The language has a limited set of commands (about eleven) and is more suitable for recreational diving as it does not provide messages to communicate an emergency and also requires the use of both hands for communication, which during a professional dive is no small constraint. The same year, Fulton et al. [39,40] proposed an interesting approach for underwater robot-to-human communication via robot motion: every motion, called a kineme, is associated with a distinct meaning (i.e., head shake (yaw) = no).

A public dataset of more than 10,000 images [12] based on the work of Chiarella et al. [10] was also released in 2019, while, in 2020, Martija et al. published a paper on this same dataset showing some possible benchmarks for both classical machine learning and deep

learning approaches [21]. In 2020, Zahn developed a single-camera system for detecting gloveless gestures underwater, testing it on nine gestures [41].

Similar to the 2019 work, in 2022, Fulton et al. [42] developed a robot-to-human language that uses light from an array of LEDs to communicate information. Each piece of information is communicated through what is called a luceme (sequences of colored light with semantic meaning). In 2023, Zhang et al. [43] used the CADDY dataset to test their image–text matching system for AUV. They proposed a visual–textual framework for underwater gesture recognition using text semantic information.

As we have seen, the existing research has mainly concentrated on detecting divers and their hand gestures themselves, rather than developing a language based on those gestures and the associated framework. As a result, the question of how to create a gesture-based language and its accompanying framework remains partially unanswered in the literature.

## 3. Methodology

The methodology used in the study involved the following elements: First, the existing language of underwater human–robot interaction, called Caddian, was revised based on feedback from divers. Subsequently, the language was expanded to include a new section that enabled collaboration and coordination among autonomous underwater vehicles (AUVs). This new extension enabled the formation of AUV groups, where actions can be allocated among them and the AUV can become the leader of the entire AUV fleet or a specific subgroup, allowing a human operator to delegate a robot leader to instruct the other robotic team members.

Following the previous study on user feedback [10], in this study the focus was redirected towards producing useful material for analysing the language and testing it on different robotic platforms. A synthetic corpus was generated by randomly generating 50 million sentences in the newly expanded Caddian language. This synthetic corpus served multiple purposes. Firstly, it allowed for exploration of the language's structure and capabilities, providing insights into the potential range of messages that can be conveyed. Secondly, it served as a benchmark for evaluating the performance of natural language processing algorithms and underwater human–robot interaction systems. The analysis conducted on the synthetic corpus revealed that, with a restricted set of gestures/lemmas, 70% of the corpus occurrences were covered, demonstrating the efficiency of the new Caddian expansion.

In the final part, to compare Caddian with other languages, a comparison framework was created, as there was no existing method for such comparisons in underwater human–robot interaction. Utilising this framework, all existing underwater human–robot interaction languages, to the best of the author's knowledge, were analyzed and compared. The results of this analysis provide a comprehensive overview of the different UHRI frameworks and highlight areas for future research and improvement in underwater human–robot interaction.

## 4. Human–Robot Communication

### *4.1. Language Update and Revision*

In this paper, the language presented in [10] has been updated and expanded based on feedback from the diving community and results from trials, which inspired the update and expansion of the language as outlined below. Specifically, divers indicated that a finite set of Caddian gestures needed to be easy and quick to perform in order to communicate important states, such as dizziness, hypoxia, and nitrogen narcosis. As a result, a subset of gestures was identified and led to the introduction of the production <slang> and the elimination of the previous productions <problem> and <p_action>. Some gestures that were identified as slang were not included in the <slang> production, but could still be used since the previous version of the language allowed them (for example, the "OK" gesture was one of them). The below illustrates some examples of simplifications for the

utterances, "I have an ear problem" (1), "I am out of breath" (2), and "Something is wrong [environment]" (3), from the old version of the language to the new one. It is evident that, in the new form, priority was given to simplicity and immediacy by constituting the new messages by one gesture/lemma instead of multiple ones.

For the benefit of the reader, we provide some guidelines to better understand the examples that follow. In the previous version of Caddian, the "ƀ" symbol was utilised to indicate the presence of a problem, which was then further specified in subsequent symbol/lemmas. Regarding other symbols employed as lemmas, the mapping between symbol and lemma in most cases was one-to-one, i.e., an alphabet symbol represented a lemma. The rationale was to select the initial letter of the corresponding word that conveyed its intended meaning. In cases where multiple symbols shared the same letter, they were distinguished by the use of subscripts. For instance, "$H_1$" denoted "Have", as the unmodified "H" had already been allocated to represent the lemma "Here". Similarly, the letter "B" without a subscript indicated "Boat", whereas, with a subscript of 2, it represented "Be out of", and, with a subscript of 3, it indicated "Breath". Lastly, the symbol "$P_g$" was designated to represent the concept of a "General problem". This symbol/lemma was used when the diver experienced a sense of discomfort or unease without being able to pinpoint the specific cause or source of potential danger.

$$A\,ƀ\,H_1\,E\,\forall \longmapsto A\,ear\,\forall \tag{1}$$

$$A\,ƀ\,B_2\,B_3\,\forall \longmapsto A\,out\_of\_breath\,\forall \tag{2}$$

$$A\,ƀ\,P_g\,\forall \longmapsto A\,prob\_gen\,\forall \tag{3}$$

The "slang" commands are listed in Table 1 and some are shown in Figure 2.

**Table 1.** Slang messages added.

| **Slang Messages** | | |
|---|---|---|
| out_of_breath = breath | K = cramp | V = vertigo |
| ear = ear | b = Something is wrong [on me] | cold = cold |
| prob_gen = generic problem (danger) | Reserve = on reserve (50 bar left) | out_of_air = to be out of air |
| low = low on air | | |
| + = up | − = down | const = stay at this depth |
| Ok = ok | No = no | U = don't understand |
| I = I, me | Y = you | boat = boat |
| Turn = Turn of 180° degrees | | |

**(a)** **(b)** **(c)** **(d)**

**Figure 2.** Examples of CADDIAN slang gestures.From left to right (**a**) "boat" (**b**) "cold" (**c**) "generic problem: danger" (**d**) "Ok".

The Caddian language, which is based on a context-free grammar (CFG), is a specialised language used for communication between divers and underwater robots (i.e., AUVs); consequently, the messages and commands defined in the language are context-dependent

and were confirmed after trials, with seven additional messages/commands added. As described in [10], a semantic function is used to map the language's gesture sequences and written forms to messages and commands. This allows gestures and written forms to be changed and associated with different interpretations since the language is agnostic to machine perception [23]. The language allows for a high degree of freedom as different back-ends can be interfaced with different front-ends.

A summary of the aforementioned changes can be found in Table 2.

**Table 2.** List of commands.

| Group | Commands/Messages | |
|---|---|---|
| **Problems** | 9 messages/commands Confirmed | |
| **Movement** | 13 messages/commands Confirmed | |
| **Interrupt** | 4 messages/commands Confirmed | |
| **Setting variables** | 9 messages/commands Confirmed | |
| **Feedback** | 3 messages/commands Confirmed | |
| **Works** | 9 messages/commands Confirmed | Turn of 180° degrees |
| **Questions (new Group)** | Where is the boat? | How much air do you have left? |
| | Are you ok? | Is there any danger? |
| **Status (new Group)** | Low on air | On reserve |

The new version of the language, from here on the Caddian Core, has been described using a BNF notation, which provides a formal way to describe the syntax and structure of the language.

The BNF notation which follows includes the new commands and productions (i.e., <slang>, <questions> and <question>) and provides a clear and concise representation of the new version of the language, making it easier for users to understand and implement.

$\langle S \rangle$ ::= A $\langle a \rangle$ $\langle S \rangle$ | $\forall$

$\langle a \rangle$ ::= $\langle slang \rangle$ | $\langle agent \rangle$ $\langle m\_action \rangle$ $\langle object \rangle$ $\langle place \rangle$ | $\langle set\_variable \rangle$ | $\langle feedback \rangle$ | $\langle interrupt \rangle$ | $\langle work \rangle$ | $\langle questions \rangle$ | $\varnothing$ | $\triangle$

$\langle slang \rangle$ ::= $\langle quantity \rangle$ | out_of_air | out_of_breath | cold | boat | ƀ | prob_gen | const | ear | cramp | vertigo | U | low | reserve

$\langle agent \rangle$ ::= I | Y | W

$\langle m\_action \rangle$ ::= take | come | do | follow | go $\langle direction \rangle$ $\langle num \rangle$

$\langle direction \rangle$ ::= forward | back | left | right | up | down

$\langle object \rangle$ ::= $\langle agent \rangle$ | $\Lambda$

$\langle place \rangle$ ::= boat | P | here | $\Lambda$

$\langle feedback \rangle$ ::= ok | no | U | $\Lambda$

$\langle set\_variable \rangle$ ::= speed $\langle quantity \rangle$ | L $\langle level \rangle$ | P | light $\langle quantity \rangle$ | air $\langle quantity \rangle$

$\langle quantity \rangle$ ::= + | -

⟨*level*⟩ ::= const | limit | free

⟨*interrupt*⟩ ::= Y ⟨ *feedback* ⟩ do

⟨*work*⟩ ::= Tes ⟨*area*⟩ | Tes ⟨*place*⟩ | Fo ⟨*area*⟩ | Fo ⟨*place*⟩ | wait ⟨*num*⟩ | check | ⟨*feedback*⟩ carry | for ⟨*num*⟩ ⟨*works*⟩ end | turn | Λ

⟨*works*⟩ ::= ⟨*work*⟩ ⟨*works*⟩ | Λ

⟨*area*⟩ ::= ⟨*num*⟩ ⟨*num*⟩ | ⟨*num*⟩

⟨*num*⟩ ::= ⟨*digit*⟩ ⟨*num*⟩ | Ψ

⟨*digit*⟩ ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0

⟨*questions*⟩ ::= U ⟨*question*⟩

⟨*question*⟩ ::= boat | air | ƀ | prob_gen

In the BNF forms of Caddian Core, many terminals are self-explanatory; however, some are not. These include ƀ which stands for "I do not feel well", ∅, which stands for "Abort mission", △, which stands for "General evacuation", "U", which stands for "I don't understand", "P", which represents "point of interest", "Tes", which represents "tessellation", and "Fo", which means "photograph".

To make the reader's experience easier, a translation table is provided (Table 3), where all commands are translated into Caddian Core. For those commands that can have variable arguments, a possible example is given. This translation table is intended to provide a clear and concise representation of the language and its capabilities, helping the reader to understand the usage of the language in the context of underwater human–robot interaction.

The table includes all the new commands added to the language, as well as the new productions. It also provides a clear representation of the language's syntax, making it easy for the reader to understand the language's capabilities and the way in which it can be used to communicate with underwater robots. To make the Caddian written form more natural and intuitive, the productions "I" and "Y" are sometimes swapped with "me" and "robot", based on whether the <agent> is the subject or direct object of the verb. As can be observed from the table, all the commands can be grouped into semantic sets highlighted in bold. In this paper, only the two new groups "Questions" and "Status" are explained; for the remaining ones, please refer to previous work for a more in-depth explanation [10]:

- **Questions**: each command in the Questions set refers to the possibility of asking the AUV a question or being questioned by it.
- **Status**: each command within this set refers to the ability to answer questions asked by the AUV.

Clearly, the written form of the Caddian language presented here is based on the written form alphabet Σ (Figure 3). In fact, considering the requirement for a language to be easily taught and learned, symbols (such as the letters of the Latin alphabet) and strings of symbols (i.e., words) that can be easily written have been employed instead of ideograms representing gestures or images depicting gestures. The signs and the written alphabet are connected by a bijective mapping function which translates them from one domain to another. Figure 3 provides an illustration of this mapping function.

For a more comprehensive view of the Caddian Core language, the website http://www.caddian.eu (accessed on 6 June 2023) provides the corresponding hand gestures for each symbol in the alphabet.

Many of these gestures have been chosen from those already used by divers all over the world and are, therefore, universally recognised with their intended meanings. The remaining gestures have been selected from those that can be executed with the hands, with

an effort to choose those that are evocative and easy to remember. For example, the gesture for "Take a photo" resembles the tripod of a camera. Table 4 shows some examples of the association. This separation of gestures from the alphabet and its context-free grammar provides robustness, while leaving the implementation of the language (i.e., the choice of gestures) to individual implementations and contexts.

**Table 3.** Translation table.

| | Message/Command | Caddian |
|---|---|---|
| **Problems** | | |
| | I have an ear problem | *A* ear ∀ |
| | I am Out of breath | *A* out_of_breath ∀ |
| | I am out of air [air almost over] | *A* out_of_air ∀ |
| | Something is wrong [diver] | *A* ƀ∀ |
| | Something is wrong, danger [environment] | *A* prob_gen ∀ |
| | I'm cold | *A* cold ∀ |
| | I have a cramp | *A* cramp ∀ |
| | I have vertigo | *A* vertigo ∀ |
| **Movement** | | |
| | Take me to the boat | *A* Y take me boat ∀<br>*A* I follow Y *A* Y come boat ∀ |
| | Take me to the point of interest | *A* Y take me P ∀<br>*A* I follow Y A Y come P ∀ |
| | Return to/come X<br>$X \in point\ of\ interest, boat, here$<br>i.e., go to point of interest, boat, come here | *A* Y come P ∀<br>*A* Y come boat ∀ or *A* boat ∀<br>*A* Y come here ∀ |
| | Go X Y<br>$X \in forward, back, left, right, Up, Down$ and $Y \in \mathbb{N}$<br>$Y \in \mathbb{N}$ | *A* Y go forward n ∀, *A* Y go back n ∀<br>*A* Y go left n ∀, *A* Y go right n ∀<br>*A* Y go up n ∀, *A* Y go down n ∀<br>n ∈ <num> |
| | You lead (I follow you) | *A* I follow Y ∀ |
| | I lead (you follow me) | *A* Y follow me ∀ |
| **Interrupt** | | |
| | Stop [interruption of action] | *A* Y no do ∀ or *A* no ∀ |
| | Let's go [continue previous action] | *A* Y ok do ∀ or *A* Y do ∀ |
| | Abort mission | *A*∅∀ |
| | General evacuation | *A* △ ∀ |
| **Setting variables** | | |
| | Slow down (during a stop decrease default movement speed) | *A* speed − ∀ |
| | Accelerate (during a stop increase default movement speed) | *A* speed + ∀ |
| | Keep this level (any action is carried out at this level) | *A* level const ∀ |
| | Free level ("Keep this level" command does not apply anymore) | *A* level free ∀ |
| | Level Off (AUV cannot fall below this level, no matter what diver says: the robot interrupts any action, if the action forces him to break this rule) | *A* level limit ∀ |
| | Set point of interest henceforth any action may refer to a point of interest) | *A* P ∀ |
| | Give me light (switch on the on board lights) | *A* light + ∀ |
| | No more light (switch off the on board lights) | *A* light - ∀ |
| | Give me air (switch on the on board oxygen cylinder) | *A* air + ∀ |
| | No more air (switch off the on board oxygen cylinder) | *A* air - ∀ |
| **Feedback** | | |
| | No (answer to repetition of the list of gestures) | *A* no ∀ |
| | Ok (answer to repetition of the list of gestures) | *A* ok ∀ |
| | I don't understand (repeat please). No idea. | *A* U ∀ |

**Table 3.** *Cont.*

| | Message/Command | Caddian |
|---|---|---|
| **Works** | | |
| | Tessellation X * Y area<br>$X, Y \in \mathbb{N}$ | *A* Te n m ∀<br>*A* Te n ∀ [square] |
| | Tessellation of point of interest/boat/here | *A* Te P ∀ |
| | Photograph of X * Y area $X, Y \in \mathbb{N}$ | *A* Fo n m ∀<br>*A* Fo n ∀ [square] |
| | Photograph of point of interest/boat/here | *A* Fo P ∀ |
| | Wait n minutes $n \in \mathbb{N}$ | *A* wait n ∀ |
| | Tell me what you're doing | *A* check ∀ |
| | Carry a tool for me | *A* carry ∀ |
| | Stop carrying the tool for me [release] | *A* no carry ∀ |
| | Turn of 180° degrees | *A* turn ∀ |
| | Make a photograph of this area n*m X times<br>n,m,X ∈ ℕ | *A* for X Fo n m end ∀<br>n,m,X ∈ <num> |
| **Questions** | | |
| | Where is the boat? | *A* U boat ∀ |
| | Are you ok? | *A* U ƀ∀ |
| | Is there any danger? | *A* U prob_gen ∀ |
| | How much air do you have left? | *A* U air ∀ |
| **Status** | | |
| | Low on air | *A* low ∀ |
| | On reserve | *A* reserve ∀ |

**Table 4.** Some examples of association gesture-written form-meaning. The images presented here were acquired with the help of the Libhand library [44]. The textures were modified using the gloves worn during missions. This library was also used in the control software for the AUV at the command center (see Figure 8).

| | | | |
|---|---|---|---|
|  |  |  |  |
| Written form: 1<br>Semantics: 1 | Written form: 2<br>Semantics: number 2 | Written form: 3<br>Semantics: number 3 | Written form: 4<br>Semantics: number 4 |
|  |  |  |  |
| Written form: 5<br>Semantics: number 5 | Written form: boat<br>Semantics: boat | Written form: carry<br>Semantics: carry equipment | Written form: mosaic<br>Semantics: do a mosaic |
|  |  |  |  |
| Written form: ok<br>Semantics: confirmation | Written form: ∀<br>Semantics: End communication | Written form: A<br>Semantics: Start message | Written form: Fo<br>Semantics: Take a photo |

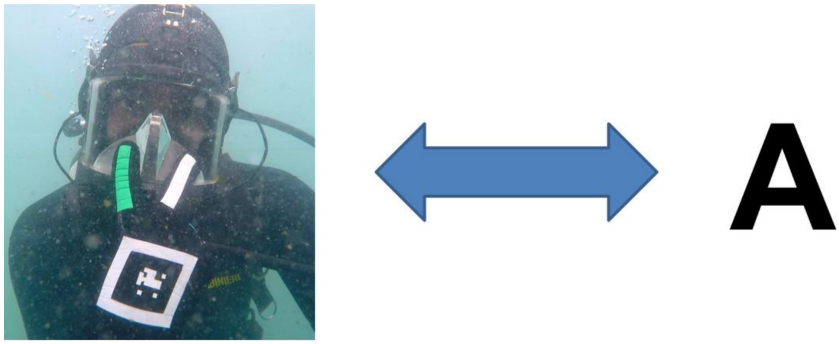


**Figure 3.** In order to write documents describing the UHRI language and its grammar and syntax, gestures have been mapped with easily writable symbols, such as letters of the Latin alphabet/numbers/words.

### *4.2. The New Framework for Multi-AUV Collaboration and Coordination*

The communication protocol was confirmed regarding the human–robot communication part. The interested reader is referred to the previous work for further details and in-depth analysis [10]. In fact, the Caddian language is used in a communication protocol that ensures error handling and strict cooperation between the diver and the AUV. The diver can query the robot at any time about the progress of a task. The AUV is equipped with three light emitters (green, orange, and red) or similar (see, for example, the background of messages on the tablet in Figure 4) to show the status of the mission (the term "mission" in the described human–robot communication refers to a series of tasks that the diver has assigned to the AUV). Green color denotes "Idle status" in which all is well, all tasks have been completed and the AUV shows that it is awaiting orders; orange color denotes "Busy status" in which all is well and the AUV indicates that it is still working on the last mission received; red color denotes "Failure status", i.e., the AUV has detected a system failure, a syntax error or issued an emergency message earlier.

The robot's mission status is crucial. Two accessibility aspects were considered: to always understand if a mission has been terminated and to be able to know the progress of a mission. The diver can interact with the AUV using commands such as "Check", "Abort mission" and "Problems". We invite the reader to read the previous article for more details [10].

The language framework presented in [9] and in [10] describes a communication scheme between a single diver and a single AUV, whereby the diver assigns missions to the AUV, as depicted in Figure 5.


**Figure 4.** The tablet has a twofold task: the first is to give feedback to the diver; the second is to show the status of the AUV to the operator at all times via background colors ( green = idle, orange = busy, red = danger/error status).

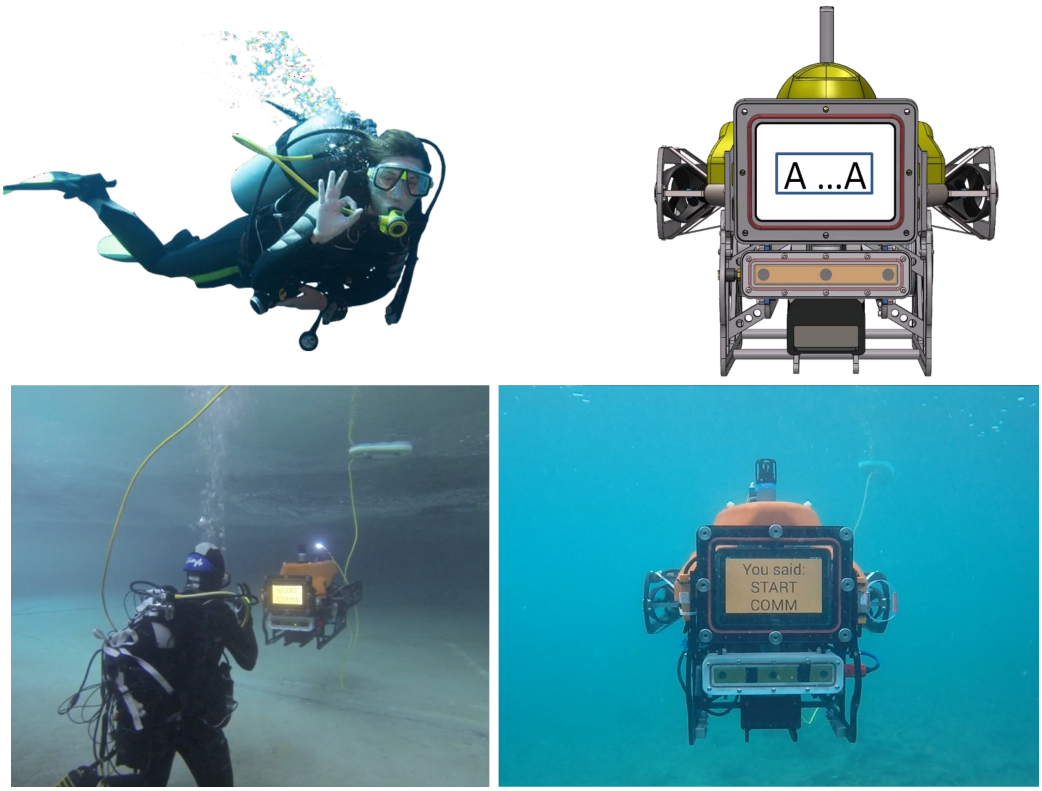


**Figure 5.** The old framework of the language envisaged only one AUV to be instructed and it was mainly a human to robot communication language. The tablet in front of the AUV is the main medium for robot to human communication.

In this article, we extend this framework to consider a scenario involving a fleet of AUVs, where the diver can instruct one or more AUVs to perform tasks or designate one AUV as a leader, which, in turn, instructs the other AUVs as subordinates.

This hierarchical structure can significantly optimise task execution time, save the batteries of AUVs and reduce risks to divers.

In the typical scenario, the diver instructs the nearest AUV, which then directs the other AUVs without the need for gathering them at a single location (see Figure 6 for a possible scenario). This approach enables hardware platform differentiation for AUV selection, as in nature, where, for example, soldier ants are more robust than workers. In the same way, the leading AUV can have a longer battery life than the subordinate AUVs so that it can reach them and provide instructions. In this way, there are two advantages: the first is that other AUVs can save the battery needed to return to base to receive new instructions and use it for other jobs, thus optimising the execution time of individual tasks and, thus, of an entire campaign of work; the second consists of the fact that this also saves the human operator time by not having to individually contact all AUVs to provide new instructions, thus reducing the risks associated with time spent diving. Moreover, the expansion of the language framework not only facilitates greater flexibility in the selection of AUV platforms, but also accommodates a variety of potential application scenarios, including, for example, those involving human-occupied vehicles and hybrid systems incorporating autonomous underwater gliders or unmanned surface support vehicles [45].

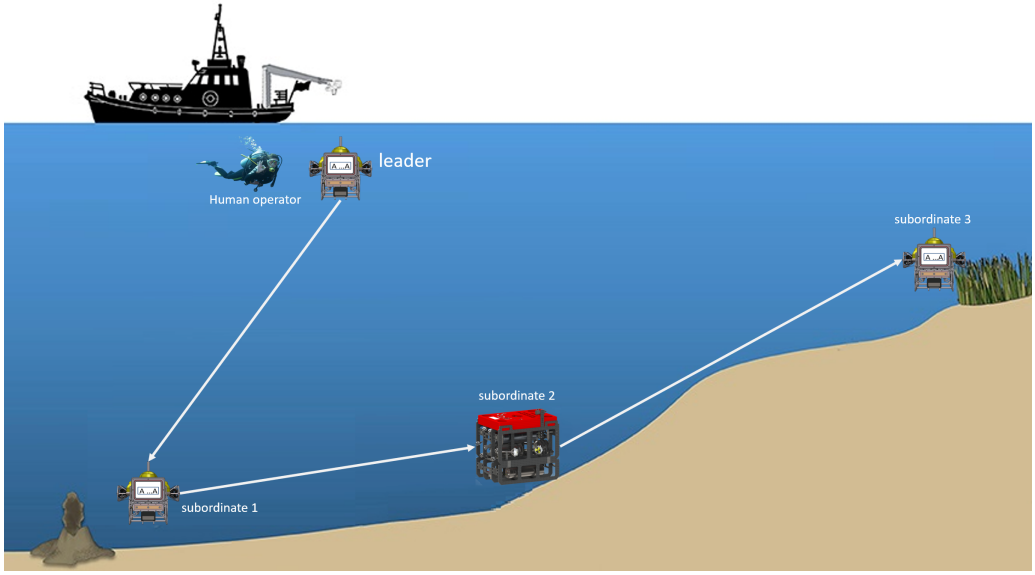


**Figure 6.** Caddian new framework: an AUV leader is instructed by a diver, then the AUV leader instructs the different AUV subordinates.

### *4.3. A Collaboration and Coordination Human–Robot Interaction Language*

In contrast to the old framework, where the AUV always had the same tasks and did not have to differentiate the actions to be taken and the commands to be received based on the context, the addition of new functionalities for coordination and collaboration requires that each AUV can have at least two collaboration states: a state in which commands are communicated to set up teamwork and hierarchy among the various robots, and a normal working state where the previous tasks that were performed in the old Caddian framework shown in previous works are carried out. In the previous framework, the AUVs were not able to distinguish between different contexts and had a limited set of tasks. This led to a lack of flexibility in adapting to different situations, which is a major challenge in underwater exploration missions. To address this issue, the new framework introduces a more complex set of tasks and commands that allow the AUVs to communicate with each other in a more sophisticated way. This results in a more efficient and adaptable system that can better handle complex missions. The two states described in this paper reflect the need for different modes of operation in the new framework (see Figure 7). The first state, defined from here on as "Group level" or "Team level", where commands are communicated to set up teamwork and a hierarchy among the AUVs, is necessary to ensure that the AUVs can work together effectively and efficiently. This state allows the AUVs to communicate and coordinate their actions, which is essential for tasks such as mapping large areas or performing complex manipulations, and, in case of the leader AUV, can enable leader tasks and functionality in both software (i.e., commands specific to the leader, such as "Starting team level" to temporarily enable the "Team level" in subordinate AUVs) and hardware (i.e., if an AUV is a leader, it can enable the secondary battery to allow it to reach other subordinate AUVs). The second state, defined from here on as "Solo level", which is the normal working state, is where the AUVs carry out the tasks that were previously performed in the old Caddian framework. These tasks may include simple navigation, data collection and diver monitoring.

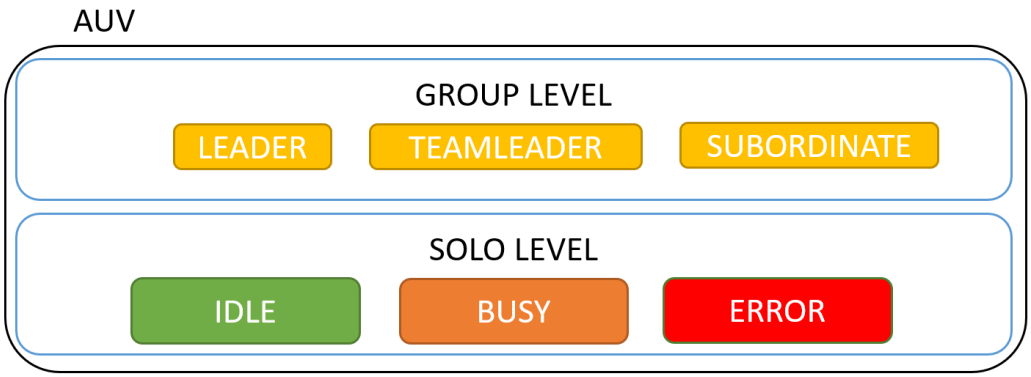


**Figure 7.** The new states introduced with the collaboration and coordination framework. In "Solo level", we can see the states described in [10], while in the "Group level", we can distinguish the three roles that an AUV can assume at the hierarchical level, namely, "Leader," "Teamleader", or "Subordinate".

#### 4.3.1. Syntax

To illustrate the framework's new features, it is useful to first introduce the syntax of the new commands available both in "Group Level" and "Solo level" mode, which is presented below. Some productions, as can be seen, are mutated from the Caddian Core language.

$\langle H \rangle$ ::= A $\langle hierarchy \rangle$ $\langle$ $H$ $\rangle$ | $\forall$

$\langle hierarchy \rangle$ ::= id $\langle num \rangle$ | $\langle mission\_h \rangle$ | $\langle mission\_order \rangle$ | $\Lambda$

$\langle mission\_h \rangle$ ::= mission $\langle leader \rangle$ $\langle teamleader \rangle$ $\langle team\_members \rangle$ | $\Lambda$

$\langle leader \rangle$ ::= $\langle num \rangle$

$\langle team\_leader \rangle$ ::= $\langle num \rangle$

$\langle team\_members \rangle$ ::= $\langle team \rangle$ $\langle team\_members \rangle$ | $\Lambda$

$\langle team \rangle$ ::= $\langle num \rangle$

$\langle mission\_order \rangle$ ::= worker $\langle num \rangle$ $\langle orders \rangle$

$\langle num \rangle$ ::= $\langle digit \rangle$ $\langle num \rangle$ | $\Psi$

$\langle digit \rangle$ ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0

$\langle orders \rangle$ ::= / $\langle order \rangle$ $\langle orders \rangle$ | $\Lambda$

$\langle order \rangle$ ::= $\langle work \rangle$ | $\langle set\_variable \rangle$ | boat | ƀ | prob_gen | $\Lambda$

In conjunction with the addition of syntax for "Team Level", the necessary commands for communication are also added in the "Solo Level". In fact, at present, the communication protocol lacks a way to display one's own identifier and make oneself recognizable as a leader, as well as a way to query an AUV to understand its identifier. As for enabling communication of the "Team level", that we can also call hierarchical communication, we start from the assumption that the leader, once it understands that it is the highest in the hierarchy (i.e., its identifier is the same as the one set through the production <leader>), automatically enables the functionality related to the ability to elicit the signal to start hierarchical communication. The signal can, for example, be a specific combination of lights, such as the ones used in previous works as a status indicator (see [9,10]), or a specific

symbol on the tablet. For a rationale of the use of light emitters, the reader can refer to the work of Fulton et al. [42].

Similarly, this applies to team leaders with the only constraint that a team leader can only give orders to their subordinates and can receive orders only from the mission leader.

In relation to this matter, it is essential to emphasise that, within this novel framework, autonomous underwater vehicles (AUVs) function in their default "Solo level" state and revert to this state upon the completion of a "Group level" communication (i.e., upon the emission of an end communication signal "∀" and the successful acknowledgment of the message by the receiving AUV). The state transition from "Solo level" to "Group level" can be accomplished by leveraging the wake word technique. In the case of inter-AUV communication, this mechanism can manifest through the activation of a specific combination of lights or the display of a designated symbol on the AUV tablet. Conversely, for the initiation of the "Group level" mode by the human operator, various modalities can be utilised, such as an activation gesture, a sequence of flashlight activation and deactivation, the utilization of an ARTag [33], or similar concepts.

That being said, the modifications to the "Solo level" productions are minimal and affect only two productions, as follows.

$\langle feedback \rangle$ ::= ok | no | U | id $\langle num \rangle$ | $\Lambda$

$\langle question \rangle$ ::= boat | air | ƀ | prob_gen | id

As can be seen from the two productions, the changes relate to the addition of the ability for the operator or AUV to ask for an AUV's ID and for the AUV alone to respond by stating its ID. Although the additions in terms of terminals are small (i.e., three new gestures for "id", "mission" and "worker") the complexity of messages compared to Caddian Core is greater; however, we would like to point out that, since messages occur between AUVs, except in the first phase of instruction of the AUV leader that could happen also through a human operator, other means of communication than gestures can be employed. In our case, the tablet, which was initially used only as a means of feedback to the diver (see Figures 4 and 8 for example), can be used in robot-to-robot communication as a means of issuing commands using the written form of Caddian accordingly.

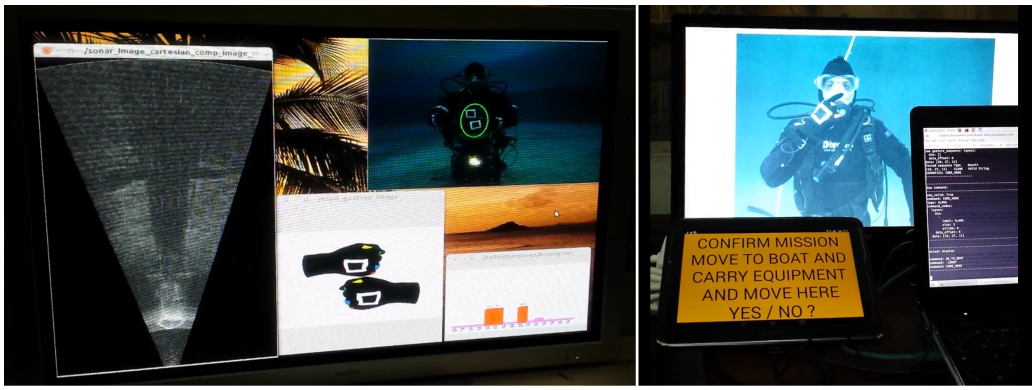


**Figure 8.** The control console used during trials that can also be replicated in the new framework.The left figure illustrates the sonar output, the gesture recognised by the classifier, and the camera image. The bottom right section of the figure exhibits a histogram depicting potential gesture categories to which the recognised gesture could belong. On the right side, the figure showcases the tablet interface displaying the request for feedback provided by the AUV to the diver. Additionally, the screen in the background depicts the diver confirming the mission, reflecting the visual perspective of the AUV.

#### 4.3.2. Semantics

The new functionalities involve a strict communication protocol that must be followed by the diver operator and AUVs and that increases the complexity of the AUV's "Mission controller". In fact, after enabling hierarchical communication (i.e., "Group level" enabled) the operator or AUV leader can (see Table 5 for examples):

- set the identifier of the AUV to which he or she is speaking;
- describe the hierarchy of an individual team—in the presence of multiple teams, the description of each individual team has to be issued with a distinct command;
- describe the tasks for an AUV that can then be assigned later to the latter after identifying it.

Regarding the last point, during communication in the "Solo level" state, the AUV can query another AUV to understand its identifier, check if it has orders for that identifier, and then switch to "Group level" to communicate them to it (this procedure is implemented at the mission controller level).

By setting up this framework, it becomes apparent that there are some possible errors arising from the fact that AUVs are not synchronised and, thus, may have inconsistency problems regarding the tasks to be performed by each individual AUV (i.e., reconfiguring tasks for a subordinate) and the hierarchical structure between them (i.e., reconfiguring a team leader only for a worker). All these problems must be handled at the diver operator level and it must be taken as an assumption that only authorised AUVs or authorised personnel can give orders at the "Group level". In addition, in the initial instruction of the AUV leader, it is essential to pay close attention to ensure there are no intersections between team members (i.e., a worker can only belong to one team) or multi-level hierarchies (i.e., a subordinate appearing as a team leader for other teams, or, worse, a team leader appearing as a subordinate in their own team).

**Table 5.** Coordination and cooperation: examples of commands.

| Message/Command | Caddian |
|---|---|
| **"Group level" enabled, "Solo level" disabled** | |
| Set identification number n | *A* id *n* Ψ ∀<br>$n \in \mathbb{N}$ |
| Set hierarchy for a mission where AUV #1 is leader and AUV #2 is team leader of #3, #4 #5 and AUV #6 is team leader of #7, #8 | *A* mission 1 Ψ 2 Ψ 3 Ψ 4 Ψ 5 Ψ ∀<br>*A* mission 1 Ψ 6 Ψ 7 Ψ 8 Ψ ∀ |
| List of orders for AUV #3: take a picture of point of interest, tesselation of 4 m × 5 m area then return to boat | *A* worker 3 Ψ /Fo P /Te 4 Ψ 5 Ψ /boat ∀ |
| **"Group level" disabled, "Solo level" enabled** | |
| Ask for identification number | *A* U id ∀ |
| Stating identification number (answer) | *A* id <num> ∀ |

## 5. Experiments and Evaluation of the Language

We discuss below the experiments conducted on the new version of the Caddian Core language and present the results of an evaluation of the new framework (Caddian Core + Collaborative and hierarchical part) with respect to other existing underwater human–robot interaction languages.

### *5.1. The Caddian Corpus*

#### 5.1.1. Experimental Setup

In previous work on Caddian, our focus was on users' understanding of language. Tests were performed on land on 22 volunteers who were given a questionnaire after receiving a small introduction to the language and its gestures. The modifications made to the language in this work simplify the syntax and use the same gestures that were used in the above tests; consequently, our attention in this work is focused on a more purely linguistic aspect of the language.

In this regard, a generator of syntactically and semantically correct Caddian sentences was created and, with it, a corpus dedicated to Caddian. The generator creates random sentences of random length between 1 and 9 (the limit of 9 is hardcoded, but can be changed in the code available, as well as the corpus, at these links https://github.com/drchiarre/Caddian_corpus (accessed on 6 June 2023), http://www.caddian.eu (accessed on 6 June 2023).

The corpus currently consists of one hundred million Caddian sentences composed of 1,645,405,039 lemmas. The creation of the corpus was performed by running the Caddian sentence generator for 100 runs, asking for each run to generate one million sentences. The length of the sentences, i.e., the number of commands they contain, was created randomly according to a normal distribution. As mentioned above, the generator creates syntactically and semantically correct sentences; however, it does not ensure pragmatic correctness (i.e., the individual commands are syntactically correct and have meaning, but it is not checked that in the order of the issue adapted to the context they make sense). While it is undeniable that an artificially generated corpus cannot match the utility of a corpus based on actual diver gestures/messages, it is important to note that, to the best of our knowledge, there are currently no existing corpora specifically focused on diver to diver or AUV to diver communication in underwater human–robot interaction (U-HRI) scenarios. In light of this limitation, a synthetic corpus holds its own value for several reasons. Firstly, the synthetic corpus allows us to delve into the structure and capabilities of the language itself, providing invaluable insights into the potential range of messages that can be effectively conveyed in Caddian. Secondly, the corpus serves as an essential benchmark for evaluating the performance of natural language processing algorithms and underwater human–robot interaction systems. By applying these algorithms to the corpus, we can assess their effectiveness in comprehending and generating responses in the context of underwater environments. This evaluation process enables us to refine and improve the algorithms, ensuring more accurate and contextually appropriate language processing. Lastly, the corpus plays a crucial role in facilitating the development of training data for machine learning models. By utilising the corpus, we can train these models to recognise and interpret Caddian language, contributing to the overall progress and effectiveness of the underwater robotic system.

#### 5.1.2. Evaluation of the Caddian Corpus

The Caddian corpus consists of one hundred files each containing a total of fifty million sentences, 499,987,270 commands and 1,645,405,039 lemmas/tokens. The average length of sentences is 4.99 commands per sentence, with an average variance of 6.66 and a standard deviation of 2.58; sentences contain from a minimum of one command to a maximum of nine commands. The commands belong to the entire domain of possible syntactically and semantically correct commands of Caddian Core.

Their distribution by files can be seen in Figure 9. From the figure, we can infer that the distribution of commands is uniform across files and that individual files share a common language pattern, one being representative of the other. According to the results and the Pareto curve presented in Figure 10, the 70% language cover threshold is surpassed after considering 15 specific gestures/lemmas, namely “start_comm”, “num_delimiter”, “Y”, “end_comm”, “P”, “boat”, “Fo”, “no”, “do”, “come”, “follow”, “carry”, “take”, “Tes”, and “me” (“start_comm” = “A”, “num_delimiter” = “∀” and “end_comm” = Ψ). This small set of gestures/lemmas accounts for approximately 70% of the occurrences in the Caddian corpus.

Consequently, by focusing on these 15 gestures for classifier training (i.e., by providing more intensive training on them), we could cover a significant portion of the language, thereby reducing the error rate and minimising misclassification.

The obtained results can be compared with those extracted from [12], as illustrated in Figure 11. The dataset used in [12] comprises real-world data collected during various trials involving the previous version of the Caddian language, which consisted of 40 messages/commands. Despite containing only 15 gestures in the dataset, which represents a subset of the total gestures in that version of Caddian, it is noteworthy that even within this dataset, the Pareto curve surpasses the 70% threshold after considering only 6 gestures/commands. Importantly, all six lemmas identified in this subset of gestures can also be found in the set of relevant lemmas presented in Figure 10.

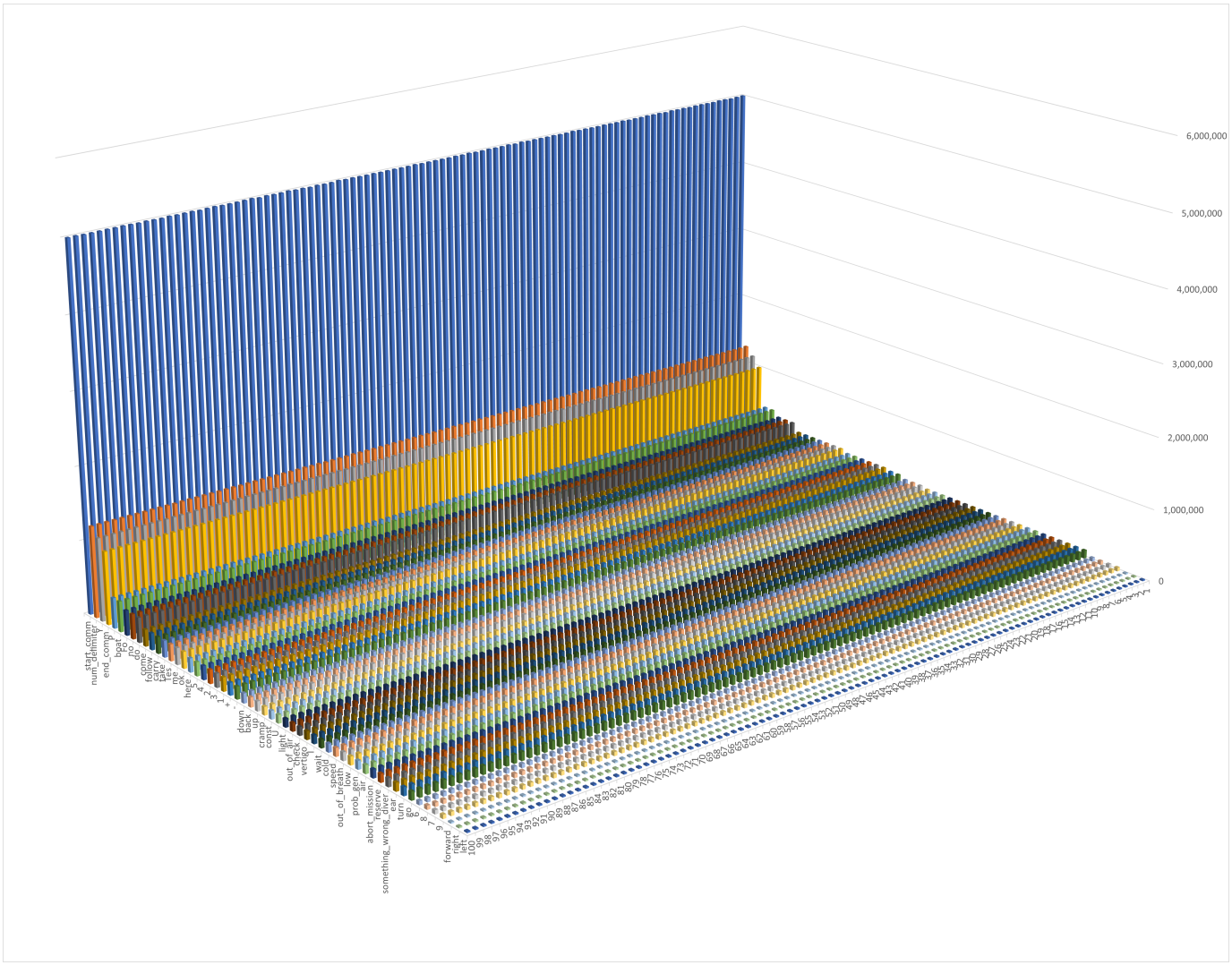


**Figure 9.** Caddian corpus: command occurrences per file. As can be seen, the command occurrences per file are quite similar in each file, although they were generated randomly.

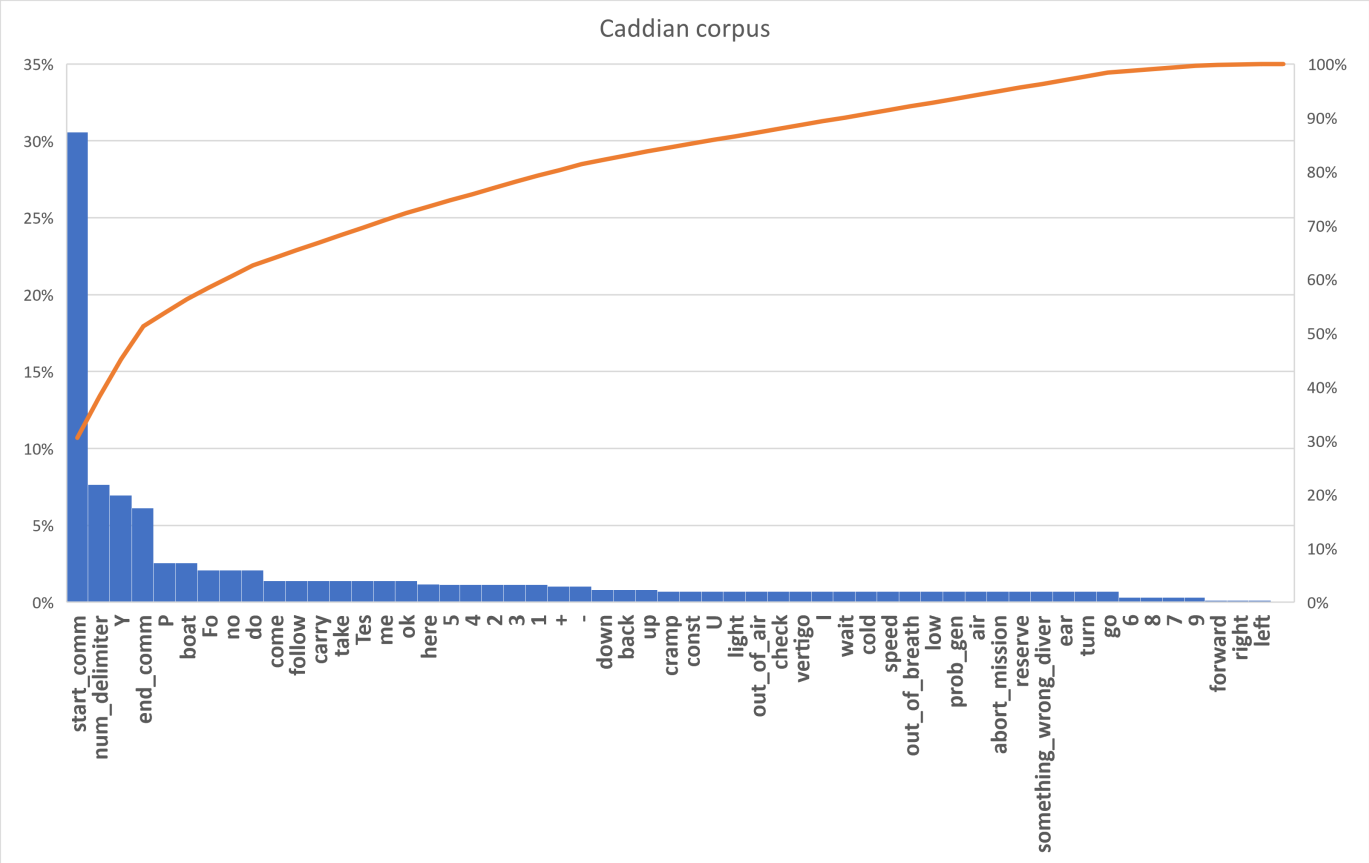


**Figure 10.** Caddian corpus: mean percentages per command and Pareto line.

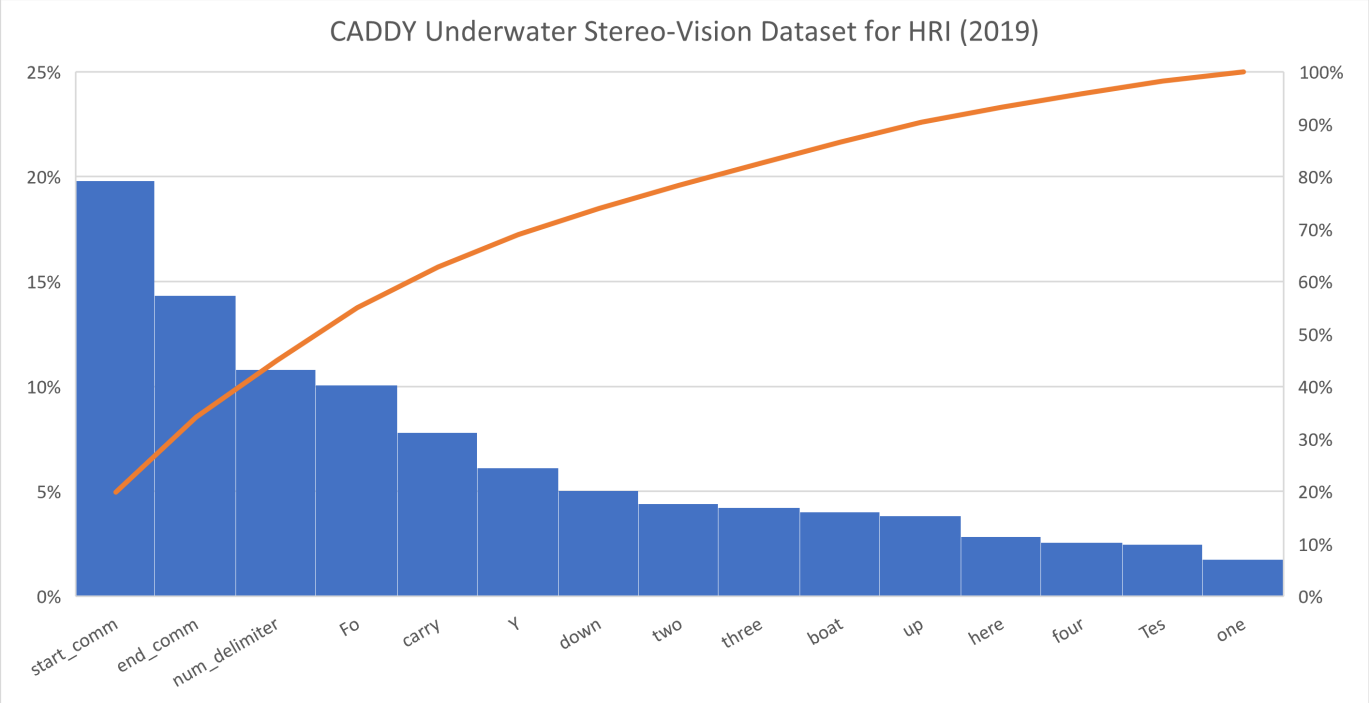


**Figure 11.** CADDY Underwater Stereo-Vision Dataset for HRI (2019) : percentages per command and Pareto line.

### 5.2. The ECU Framework and Language Comparison

#### 5.2.1. Evaluation Criteria

In order to perform an evaluation of the language that could be compared with other existing UHRI languages, we wanted to define an evaluation framework, called ECU (evaluation criteria for UHRI), with the dual purpose of ensuring, on the one hand, an objective evaluation, and, on the other hand, identifying the languages' strengths, weaknesses and related areas for improvement and, thus, help the development of new languages and frameworks that overcome the limitations of existing ones.

The comparison of the new expanded and updated version of Caddian considered the following parameters that are part of the ECU framework:

- Context-free grammar: whether or not the language has a context-free grammar. A context-free grammar is structured in such a way as to make it easier and more intuitive to understand the grammatical rules by removing some of the ambiguities that may be present in other forms of grammar. The ease of analysis which this entails affects language learning, which it represents, since learning becomes a matter of recognising individual units of syntax and the rules for combining them. In addition, a CFG has the advantage of simplicity of representation—it can be represented in a standardised form, such as the BNF, which simplifies the description of the grammar and facilitates its implementation in a language analysis system, making its programming and the creation of syntactic control rules more efficient. Both these factors make it possible to create more efficient and accurate syntactic analysis tools. Finally, it provides expandability—new syntactic constructs or terms can be added to the language without having to completely redesign the grammar. This is possible because a context-free grammar is based on defining syntactic rules that specify how different parts of speech combine to create grammatically correct sentences. These rules can be modified or extended to accommodate new syntactic constructions or new words.
- Ease of learning the language and user interface: this feature describes in language testing if the ease of learning the language was evaluated.
- Feedback: this feature indicates whether the language has included communication feedback from the robot to the human operator. Effective feedback enables human operators to better understand the state of the system, correctly perform the actions necessary to complete an assigned task, and improve the trust and safety of human operators. In fact, feedback can also affect human operators' perceptions of the reliability and quality of systems, and can, therefore, be an important factor in choosing which systems to use in certain applications.
- Single task execution speed: this feature describes whether in testing the language framework how task execution speed was evaluated.
- Task execution accuracy: this feature describes whether the accuracy of the responses provided by the robotic systems was evaluated in the testing of the language framework.
- Adaptability: this feature describes whether the ability of the systems to adapt to unforeseen or unexpected situations, such as sudden changes in environmental conditions, was predicted in the language framework. It refers to the construction of conditional statements, such as "If the visibility is poor, then wait for five minutes", or "If the current starts to increase, then return to the base". These conditional statements serve as examples of how the language framework can incorporate environmental conditions and adapt the robot's behavior accordingly.
- Robustness: in testing the language, the ability of the systems to maintain performance even in the presence of hardware or software errors or failures, such as the loss of one or more sensors, was evaluated.
- Parameters reconfiguration: the language allows reconfiguration of mission parameters. In an underwater human robot interaction context, it is important to provide for the ability to reconfigure mission parameters because environmental conditions

and mission requirements may change over time. For example, the navigational depth of an underwater robot may need to be changed due to an unexpected current or an unexpected obstacle. The ability to reconfigure mission parameters in real time allows robots to adapt to changing environmental conditions and perform their tasks more efficiently and accurately. In addition, reconfiguring mission parameters can enable robots to optimise their performance and save energy, thereby increasing mission autonomy.

- Gesture based: the language is gesture based or not. The other medium used is indicated if this was not the case. The use of gestures, in fact, can help overcome difficulties in submarine communication where audio and voice are limited. Gestures can be easily visible and recognizable even in unclear water or low light. In addition to this, gesture-based communication, which has already been adopted by the underwater community, can reduce the fatigue and stress of human personnel interacting with underwater robots, as it does not require the use of voice devices or other tools that may be cumbersome or impractical in the underwater environment.
- Human to robot: language enables human–robot communication. It is the pivotal feature of the chosen languages, but it is shown for completeness.
- Robot to human: language enables robot–human communication. In an underwater human robot interaction language, the robot may have to provide important information to the human, such as the conditions of the surrounding environment. In addition, the robot may have to report any problems or malfunctions to the human. Thus, effective and well-structured two-way communication can improve collaboration between human and robot, increasing the efficiency and safety of operations performed in the underwater environment.
- Collaboration: the language allows or provides for collaboration between AUVs. In some application scenarios, multiple AUVs may need to be used simultaneously to perform complex tasks. For example, a team of AUVs could be used to conduct detailed mapping of a marine area, where each vehicle has a specific task or, in the case of environmental emergencies, such as oil spills or marine pollution, a team of AUVs can be used to assess the size and extent of damage, map the affected area, and support recovery activities. In this case, effective communication between vehicles can improve the efficiency of the work performed. In addition to this, in an underwater environment, environmental conditions can be changeable and unpredictable, and a fleet of AUVs can collaborate to manage these situations. In addition, collaboration between AUVs can improve the efficiency of work performed and increase the safety of operations. Effective communication between vehicles can enable optimal task distribution and sharing of collected information, avoiding duplication of effort and maximising coverage of the marine area of interest.
- Scalability: in testing the framework, the ability of the robotic systems to operate effectively and efficiently, even when the number of AUVs involved increases, was evaluated.
- Overall efficiency: this criterion checks how long it took the systems to perform a specific task, such as how long it took a group of AUVs to complete an assigned task.
- Hierarchical organization: the language allows a hierarchical organization of AUVs and consequently of their tasks. This factor is very important for coordination of activities—when there are multiple AUVs operating simultaneously, it is important to have a hierarchical system to coordinate activities effectively and safely. This helps to avoid collisions and conflicts of activities. The ability to set a hierarchical order is, in addition, essential for resource management because it allows for efficient management of available resources, such as battery life, payload capacity, and availability of specific sensors. It simultaneously ensures flexibility and adaptability; the hierarchy of AUVs makes it easy to adapt activities according to terrain conditions, operator demands, or resource availability. For example, a lower-level hardware AUV could be

tasked to monitor a specific area, while a higher-level hardware AUV could be tasked to perform a more complex search mission.

- "Open sea trials" or "Pool or closed water testing": language was tested in open sea or only in pool or closed water. Open water trials are as crucial as pool or closed-water trials for the following reasons:
  - Real-world conditions and test validity: open-water testing allows underwater robotic systems to be tested under real-world conditions, similar to those that may be encountered during real-world underwater operations. This can provide important information about the robot's performance, capabilities, and limitations under real-world conditions.
  - Complex environment and variance of conditions: the marine environment is complex and dynamic, with currents, waves, and other environmental variables that can affect robot performance. Conducting trials in the open ocean allows for testing the ability of underwater robots to adapt to a wide range of conditions, improving their ability to perform real-world missions in diverse conditions.
  - human–robot collaboration: open-water trials allow testing of human–robot collaboration in a real-world environment, helping to identify any challenges or communication issues that may not emerge during pool or closed-water trials.
- Community feedback: the language has received feedback from end-users such as professional and amateur divers. It is important to have a part of the language of underwater human robot interaction derived from feedback from the community of professional and amateur divers for several reasons:
  - Domain knowledge: the community of professional and amateur divers has in-depth knowledge of the underwater domain and the activities that are performed in this environment. Incorporating their feedback into the language of underwater human robot interaction helps to ensure that the language is appropriate for the context and takes into account important aspects that might be overlooked by those who are not experts in scuba diving.
  - Usability: incorporating feedback from professional and amateur divers can improve the usability of underwater robotic systems. Divers can provide useful information about the ease of use of underwater robotic systems, their training requirements, and their ability to meet users' needs.
  - Safety: professional and amateur divers have a thorough understanding of the risks and challenges that can be encountered in an underwater environment. Incorporating their feedback can help improve the safety of underwater robotic systems, helping to prevent accidents and to ensure that underwater robots operate safely and responsibly.
- Open data: the language framework makes databases of examples or data available to the scientific community, which is useful for reproducing experiments and improving the framework. In addition to these two elements, open data are critical for knowledge sharing that can lead to greater collaboration among researchers and the discovery of new techniques and ideas.

#### 5.2.2. Comparison with Existing U-HRI Languages

As already seen in the state of the art, there is only a small number of UHRI languages and many of the works rightly focus more on solving the correct identification of gestures or the communication medium in general rather than defining a complete communication framework. In this section, we aim to compare the different frameworks for underwater human–robot interaction, focusing on the ECU parameters (context-free grammar, ease of learning the language and user interface, feedback, single task execution speed, task execution accuracy, adaptability, robustness, parameters reconfiguration, gesture-based, human to robot, robot to human, collaboration, scalability, overall efficiency, hierarchical

organization, open sea trials, pool or closed-water testing, community feedback, and open data), and identify any missing criteria. We compare several frameworks with the one presented in this article. The frameworks identified are based on the following works:

1. A Visual Language for Robot Control and Programming: A Human-Interface Study (2007) [33]
2. Gesture-based Language for Diver-Robot Underwater Interaction ( Caddian 2015) [9]
3. A Novel Gesture-Based Language for Underwater Human–Robot Interaction [10], Underwater Stereo-Vision Dataset for Human–Robot Interaction (HRI) in the Context of Diver Activities [12] and Underwater Vision-Based Gesture Recognition: A Robustness Validation for Safe Human–Robot Interaction [22] (Caddian 2018–2021)
4. Dynamic Reconfiguration of Mission Parameters in Underwater Human–Robot Collaboration [46] and Understanding Human Motion and Gestures for Underwater Human–Robot Collaboration [35] (2018–2019)
5. Robot Communication Via Motion: A Study on Modalities for Robot-to-Human Communication in the Field [39] (2019–2022)
6. HREyes: Design, Development, and Evaluation of a Novel Method for AUVs to Communicate Information and Gaze Direction [42] (2022)
7. This work (2023).

Table 6 shows the different frameworks and their corresponding values for the ECU parameters: the values zero and one indicate false and true, respectively; for the gesture-based parameter, the different mediums used for communication are indicated if other than gestures. Notably, frameworks 5 and 6 focus on robot to human communication and seek to identify communication units based on movement (kineme) and light (luceme), respectively. Frameworks 1 and 4 focus on human to robot communication, while frameworks 2, 3 and 7 identify the different versions of Caddian and describe the evolution of the framework through the revision of some parts and the addition of others to offer more functionality.

**Table 6.** Comparison of existing HRI frameworks using ECU.

| **ECU** | **Language Frameworks** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Framework Parameters** | #1 | #2 | #3 | #4 | #5 | #6 | #7 |
| Context-free grammar | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Ease of learning the language and user interface | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Feedback | 0 | 1 | 1 | 0 | N/A | N/A | 1 |
| Single-task execution speed | 1 | 0 | 0 | 1 | N/A | N/A | 0 |
| Task execution accuracy | 1 | 1 | 1 | 1 | N/A | N/A | 1 |
| Adaptability | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Robustness | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Parameters reconfiguration | 0 | 1 | 1 | 1 | N/A | N/A | 1 |
| Gesture-based | ARTag | 1 | 1 | 1 | kineme | lukeme | 1 |
| Human to robot | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Robot to human | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| Collaboration | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Scalability | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Overall efficiency | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hierarchical organization | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Open sea trials | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| Pool or closed-water testing | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Community feedback | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| Open data | N/M [1] | 1 | 1 | N/M [1] | N/M [1] | N/M [1] | 1 |

[1] To the best of the author's knowledge not mentioned in articles describing the framework.

From the table, we can conclude that none of the frameworks satisfies all the parameters, and that several frameworks can complement each other to build a complete communication framework that meets all the ECU requirements.

## 6. Challenges and Mitigation Strategies

The creation of the underwater human–robot interaction (UHRI) language and its subsequent expansion involved several challenges, which required careful consideration and actions to address them effectively. During the development of the language, one of the main challenges was to design a language that could be easily understood and used by both humans and robots. This involved simplifying the language and making it intuitive, especially for non-technical users. To overcome this challenge, we adopted already used gestures from the divers' community, we undertook extensive user testing, and iterative design processes were carried out, taking into account feedback from experts and potential users. The language was refined based on this feedback, ensuring its usability and comprehensibility. Field trials were conducted to validate and assess the performance of the UHRI language in real-world underwater environments. Challenges encountered during the field trials included environmental factors, such as poor visibility and varying conditions, which affected the effectiveness of communication between humans and robots. Pre-dive training of the divers also required careful planning, and communication once submerged proved to be a time-wasting problem. Consequently, a protocol also had to be developed to notify the divers if the trials of the tests had to be interrupted to be repeated. Feedback from divers played a crucial role in refining the UHRI language. Their input highlighted the specific challenges faced in underwater environments and provided insights into the practicality and effectiveness of the language. This feedback was carefully analyzed and incorporated into the language design and revisions, ensuring that the resulting language met the needs and expectations of the divers. The creation of the new part on the language for AUV collaboration and coordination, along with the introduction of AUV leaders and sub-leaders, as well as the transition from the "Group level" to the "Solo level", posed several challenges during the design process. One of the primary challenges was designing a language that effectively conveyed instructions for AUV collaboration and coordination. This required careful consideration of the syntax, vocabulary, and gestures used in the language to ensure clear and unambiguous communication, in particular the part where "per AUV" mission can be defined (i.e., productions <hierarchy> and <mission_order>). Creating the synthetic corpus for testing and evaluation purposes presented its own challenges. Generating 50 million sentences in the Caddian language required careful consideration of the language's grammar, vocabulary, and syntactic structures. The challenge here was to ensure that the synthetic corpus represented a diverse range of possible messages, providing a comprehensive evaluation resource for the language. The development of the evaluation framework for UHRI systems also posed challenges. As there was no existing method for comparing different UHRI frameworks, a new framework had to be devised. This involved identifying key parameters and criteria to evaluate the performance and effectiveness of the systems. The challenge was to create a framework that captured the essential aspects of UHRI and enabled meaningful comparisons between different systems. Overall, addressing these challenges involved a combination of iterative design processes, user feedback incorporation, adaptability features and careful corpus generation.

## 7. Conclusions and Future Work

In this paper, we presented a new version of a UHRI language along with its new framework. This framework enables collaboration among multiple robots to perform complex tasks. In addition, the framework allows for organising work in teams and setting a hierarchical order among AUVs. This study, as it pertains to the new version of the language, has shown, through the creation and analysis of a corpus consisting of fifty million Caddian missions, totaling about 500 million commands, that, by focusing the training of a classifier on one-third of the language gestures, a coverage of more than 70 percent

of the language is achieved. On the other hand, we recognise the constraints associated with utilising an artificially generated corpus. Consequently, we envision future endeavors aimed at addressing these limitations through more comprehensive experiments and trials, specifically focusing on the collection of diver–diver and diver–AUV communication data. By incorporating real-world interactions, we aim to refine the Caddian language and establish a comparative analysis between its usage patterns and authentic diver utterances and practical use cases. While we acknowledge that these datasets may possess certain constraints concerning naturalness due to their experimental nature, we firmly believe that they will still provide invaluable insights into the efficacy and frequency of Caddian language implementation within underwater human–robot interaction scenarios. These future works are crucial steps towards enhancing the authenticity and applicability of our research findings. Regarding the corpus creation, it is important to acknowledge that the automatic creation of semantically coherent missions, involving sequences of actions that have meaning and consistency in the real world, remains an ongoing challenge.

As for the complete framework, on the other hand, an evaluation method has been implemented for comparison with other existing systems. The application of this method shows that there is no framework that completely satisfies all evaluation parameters, but an approach that considers various elements from different languages might be the way to go. This work is a first step and, in the light of the comparison, appears to be the most comprehensive, although it still has limitations in some areas.

These areas, that have limitations in almost all of the identified frameworks, are "adaptability," "robustness," "scalability," and "overall efficiency", respectively. Based on the above limitations, there are several potential directions for future research and development. First, to address the limitation related to adaptability, future work on the UHRI language side could focus on improving the language framework to incorporate a wider range of conditional statements that can account for a wider variety of unforeseen or unexpected environmental situations that, in parallel, also need to be taken into account in the robot's control system. Second, to improve the robustness of the system, further research could be conducted to identify and address specific failure modes or errors that are not adequately handled by the current framework and to test what failures and errors the human–robot and robot–robot interaction system is tolerant to (i.e., communication can continue to occur). This could involve implementing additional language-level error detection and recovery mechanisms, as well as conducting rigorous testing under different operating conditions to ensure system resilience at the hardware level. Third, scalability can be addressed by studying the performance of the language framework when operating with a larger number of AUVs. This could involve studying communication protocols, coordination strategies, and resource allocation mechanisms to ensure effective collaboration and efficient task allocation among multiple robots. Finally, to improve overall efficiency, future work could focus on optimising the framework's task execution algorithms and decision-making processes. This could involve exploring techniques such as task prioritization, workload balancing, and resource optimization to minimise the time required to complete assigned tasks.

It might also be appropriate, within the proposed Caddian framework, to test for robot–human communication the new media identified in the other works considered (i.e., kineme, luceme) and compare them with the tablet used in the actual one. Despite the inherent challenges, it is crucial to address the study of user interaction within real-world settings in the future, particularly in relation to the framework's hierarchical collaboration as it is applied to problem-solving scenarios in practical contexts. Such investigations will provide valuable insights into the overall effectiveness of the new framework in real-world usage scenarios. Additionally, it is important to consider conducting further user satisfaction testing for the expanded framework to gain insights into which components are well-received and which may require improvement. In fact, understanding the users' preferences and opinions will be instrumental in refining and optimising the framework to ensure its acceptance, usability and effectiveness in real-world applications.

By addressing the aforementioned limitations and pursuing the described future research directions, it will be possible to further advance the capabilities and effectiveness of frameworks for human–robot interaction underwater, ultimately enabling more robust, adaptable, scalable, and efficient collaborations between humans and robots in underwater environments.

**Funding:** The research leading to these results received initial funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 611373.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available at https://github.com/drchiarre/Caddian_corpus and http://www.caddian.eu (accessed on 6 June 2023).

**Acknowledgments:** The author would like to thank Andrea Ranieri and Paola Cutugno for their invaluable assistance and support during the writing of this manuscript.

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Denoble, P.J.; Caruso, J.L.; Dear, G.d.L.; Pieper, C.F.; Vann, R.D. Common causes of open-circuit recreational diving fatalities. *Undersea Hyperb. Med. J.* **2008**, *35*, 393–406.
2. Richardson, D. *PADI Open Water Diver Manual*; PADI: Rancho Santa Margarita, CA, USA, 2010.
3. Halstead, B. Line dancing and the buddy system. *South Pac. Underw. Med. Soc. J.* **2000**, *30*, 701–713.
4. Li, S.; Qu, W.; Liu, C.; Qiu, T.; Zhao, Z. Survey on high reliability wireless communication for underwater sensor networks. *J. Netw. Comput. Appl.* **2019**, *148*, 102446. [CrossRef]
5. CMAS Swiss Diving. Segni Convenzionali. 2003. Available online: https://www.cmas.ch/docs/it/downloads/codici-comunicazione-cmas/it-Codici-di-comunicazione-CMAS.pdf (accessed on 6 June 2023).
6. Jorge Mezcua. Diving Signs You Need to Know. HTML Page. 2012. Available online: http://www.fordivers.com/en/blog/2013/09/12/senales-de-buceo-que-tienes-que-conocer/ (accessed on 6 June 2023).
7. Recreational Scuba Training Council. Common Hand Signals for Recreational Scuba Diving. Online PDF. 2005. Available online: http://www.neadc.org/CommonHandSignalsforScubaDiving.pdf (accessed on 6 June 2023).
8. Scuba Diving Fan Club. Most Common Diving Signals. HTML Page. 2016. Available online: http://www.scubadivingfanclub.com/Diving_Signals.html (accessed on 6 June 2023).
9. Chiarella, D.; Bibuli, M.; Bruzzone, G.; Caccia, M.; Ranieri, A.; Zereik, E.; Marconi, L.; Cutugno, P. Gesture-based language for diver-robot underwater interaction. In Proceedings of the OCEANS 2015, Genova, Italy, 18–21 May 2015; pp. 1–9. [CrossRef]
10. Chiarella, D.; Bibuli, M.; Bruzzone, G.; Caccia, M.; Ranieri, A.; Zereik, E.; Marconi, L.; Cutugno, P. A Novel Gesture-Based Language for Underwater Human–Robot Interaction. *J. Mar. Sci. Eng.* **2018**, *6*, 91. [CrossRef]
11. Chomsky, N. Three models for the description of language. *IRE Trans. Inf. Theory* **1956**, *2*, 113–124. [CrossRef]
12. Gomez Chavez, A.; Ranieri, A.; Chiarella, D.; Zereik, E.; Babić, A.; Birk, A. CADDY Underwater Stereo-Vision Dataset for Human–Robot Interaction (HRI) in the Context of Diver Activities. *J. Mar. Sci. Eng.* **2019**, *7*, 16. [CrossRef]
13. Mišković, N.; Pascoal, A.; Bibuli, M.; Caccia, M.; Neasham, J.A.; Birk, A.; Egi, M.; Grammer, K.; Marroni, A.; Vasilijević, A.; et al. CADDY project, year 3: The final validation trials. In Proceedings of the OCEANS 2017, Aberdeen, UK, 19–22 June 2017; pp. 1–5. [CrossRef]
14. Stilinović, N.; Nađ, Đ.; Mišković, N. AUV for diver assistance and safety-Design and implementation. In Proceedings of the OCEANS 2015, Genova, Italy, 18–21 May 2015; pp. 1–4. [CrossRef]
15. Mišković, N.; Pascoal, A.; Bibuli, M.; Caccia, M.; Neasham, J.A.; Birk, A.; Egi, M.; Grammer, K.; Marroni, A.; Vasilijević, A.; et al. CADDY Project, Year 1: Overview of Technological Developments and Cooperative Behaviours. *IFAC-PapersOnLine* **2015**, *48*, 125–130. [CrossRef]
16. Nađ, Đ.; Mandić, F.; Mišković, N. Using Autonomous Underwater Vehicles for Diver Tracking and Navigation Aiding. *J. Mar. Sci. Eng.* **2020**, *8*, 413. [CrossRef]
17. Odetti, A.; Bibuli, M.; Bruzzone, G.; Caccia, M.; Spirandelli, E.; Bruzzone, G. e-URoPe: A reconfgurable AUV/ROV for man-robot underwater cooperation. *IFAC-PapersOnLine* **2017**, *50*, 11203–11208. [CrossRef]
18. CADDY Underwater Stereo-Vision Dataset. Website. 2019. Available online: http://www.caddian.eu (accessed on 6 June 2023).
19. Jiang, Y.; Zhao, M.; Wang, C.; Wei, F.; Wang, K.; Qi, H. Diver's hand gesture recognition and segmentation for human–robot interaction on AUV. *Signal Image Video Process.* **2021**, *15*, 1899–1906. [CrossRef]

20. Yang, J.; Wilson, J.P.; Gupta, S. DARE: AI-based Diver Action Recognition System using Multi-Channel CNNs for AUV Supervision. *arXiv* **2020**, arXiv:2011.07713.
21. Martija, M.A.M.; Dumbrique, J.I.S.; Naval, P.C., Jr. Underwater Gesture Recognition Using Classical Computer Vision and Deep Learning Techniques. *J. Image Graph.* **2020**, *8*, 9-14. [CrossRef]
22. Gomez Chavez, A.; Ranieri, A.; Chiarella, D.; Birk, A. Underwater Vision-Based Gesture Recognition: A Robustness Validation for Safe Human–Robot Interaction. *IEEE Robot. Autom. Mag.* **2021**, *28*, 67–78 . [CrossRef]
23. Birk, A. A Survey of Underwater Human-Robot Interaction (U-HRI). *Curr. Robot. Rep.* **2022**, *3*, 199–211. [CrossRef]
24. Sattar, J.; Dudek, G. Where is your dive buddy: Tracking humans underwater using spatio-temporal features. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 3654–3659. [CrossRef]
25. Buelow, H.; Birk, A. Gesture-recognition as basis for a human robot interface (HRI) on a AUV. In Proceedings of the OCEANS'11 MTS/IEEE KONA, Waikoloa, HI, USA, 19–22 September 2011; pp. 1–9. [CrossRef]
26. DeMarco, K.J.; West, M.E.; Howard, A.M. Sonar-Based Detection and Tracking of a Diver for Underwater Human-Robot Interaction Scenarios. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 2378–2383. [CrossRef]
27. Chavez, A.G.; Pfingsthorn, M.; Birk, A.; Rendulić, I.; Mišković, N. Visual diver detection using multi-descriptor nearest-class-mean random forests in the context of underwater Human Robot Interaction (HRI). In Proceedings of the OCEANS 2015, Genova, Genova, Italy, 18–21 May 2015; pp. 1–7. [CrossRef]
28. Islam, M.J.; Sattar, J. Mixed-domain biological motion tracking for underwater human–robot interaction. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 4457–4464. [CrossRef]
29. Chavez, A.G.; Mueller, C.A.; Birk, A.; Babic, A.; Miskovic, N. Stereo-vision based diver pose estimation using LSTM recurrent neural networks for AUV navigation guidance. In Proceedings of the OCEANS 2017, Aberdeen, UK, 19–22 June 2017; pp. 1–7. [CrossRef]
30. Xia, Y.; Sattar, J. Visual Diver Recognition for Underwater Human-Robot Collaboration. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 6839–6845. [CrossRef]
31. Remmas, W.; Chemori, A.; Kruusmaa, M. Diver tracking in open waters: A low-cost approach based on visual and acoustic sensor fusion. *J. Field Robot.* **2021**, *38*, 494–508. [CrossRef]
32. Jiang, Y.; Zhao, M.; Wang, C.; Wei, F.; Hong, Q. A Method for Underwater Human–Robot Interaction Based on Gestures Tracking with Fuzzy Control. *Int. J. Fuzzy Syst.* **2021**, *23*, 2170–2181. [CrossRef]
33. Dudek, G.; Sattar, J.; Xu, A. A Visual Language for Robot Control and Programming: A Human-Interface Study. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007; pp. 2507–2513. [CrossRef]
34. Backus, J.W. The Syntax and Semantics of the Proposed International Algebraic Language of the Zurich ACM-GAMM Conference. In Proceedings of the International Conference on Information Processing, Paris, France, 15–20 June 1959.
35. Islam, M.J.; Ho, M.; Sattar, J. Understanding human motion and gestures for underwater human–robot collaboration. *J. Field Robot.* **2019**, *36*, 851–873.
36. Cuan, C.; Lee, E.; Fisher, E.; Francis, A.; Takayama, L.; Zhang, T.; Toshev, A.; Pirk, S. Gesture2Path: Imitation Learning for Gesture-aware Navigation. *arXiv* **2022**, arXiv:2209.09375.
37. Menix, M.; Miskovic, N.; Vukic, Z. Interpretation of divers' symbolic language by using hidden Markov models. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 976–981. [CrossRef]
38. Mišković, N.; Bibuli, M.; Birk, A.; Caccia, M.; Egi, M.; Grammer, K.; Marroni, A.; Neasham, J.; Pascoal, A.; Vasilijević, A.; et al. Overview of the FP7 project "CADDY—Cognitive Autonomous Diving Buddy". In Proceedings of the OCEANS 2015, Genova, Italy, 18–21 May 2015; pp. 1–5. [CrossRef]
39. Fulton, M.; Edge, C.; Sattar, J. Robot Communication Via Motion: Closing the Underwater Human-Robot Interaction Loop. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 4660–4666. [CrossRef]
40. Enan, S.S.; Fulton, M.; Sattar, J. Robotic Detection of a Human-Comprehensible Gestural Language for Underwater Multi-Human-Robot Collaboration. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; pp. 3085–3092. [CrossRef]
41. Zahn, M. Development of an underwater hand gesture recognition system. In Proceedings of the Global Oceans 2020: Singapore—U.S. Gulf Coast, Live Virtual, 5–14 October 2020; pp. 1–8 . [CrossRef]
42. Fulton, M.; Prabhu, A.; Sattar, J. HREyes: Design, Development, and Evaluation of a Novel Method for AUVs to Communicate Information and Gaze Direction. *arXiv* **2022**, arXiv:2211.02946.
43. Zhang, Y.; Jiang, Y.; Qi, H.; Zhao, M.; Wang, Y.; Wang, K.; Wei, F. An Underwater Human–Robot Interaction Using a Visual–Textual Model for Autonomous Underwater Vehicles. *Sensors* **2023**, *23*, 197. [CrossRef]
44. Šarić, M. LibHand: A Library for Hand Articulation, 2011. Version 0.9. Available online: http://www.libhand.org/ (accessed on 6 June 2023)

45. Yang, L.; Zhao, S.; Wang, X.; Shen, P.; Zhang, T. Deep-Sea Underwater Cooperative Operation of Manned/Unmanned Submersible and Surface Vehicles for Different Application Scenarios. *J. Mar. Sci. Eng.* **2022**, *10*, 909. [CrossRef]
46. Islam, M.J.; Ho, M.; Sattar, J. Dynamic Reconfiguration of Mission Parameters in Underwater Human-Robot Collaboration. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 6212–6219. [CrossRef]