**VI SPEC: A graphical tool for elicitation of MTL requirements**

Bardh Hoxha¹, Nikolaos Mavridis² and Georgios Fainekos¹

**Abstract**—One of the main barriers preventing widespread use of formal methods is the elicitation of formal specifications. Formal specifications facilitate the testing and verification process for safety critical robotic systems. However, handling the intricacies of formal languages is difficult and requires a high level of expertise in formal logics that many system developers do not have. In this work, we present a graphical tool designed for the development and visualization of formal specifications by people that do not have training in formal logic. The tool enables users to develop Specifications using a graphical formalism which is then automatically translated to Metric Temporal Logic (MTL). In order to evaluate the effectiveness of our tool, we have also designed and conducted a usability study with cohorts from the academic student community and industry. Our results indicate that both groups were able to define formal requirements with high levels of accuracy. Finally, we present applications of our tool for defining specifications for operation of robotic surgery and autonomous quadcopter safe operation.

I. INTRODUCTION

As robots become commercially available, their correct operation is of paramount importance. Especially for safety critical systems, safety must be guaranteed. As for example in autonomous vehicles [24] and medical robots [17], [13].

Safety requirements are usually expressed in natural language, which is inherently ambiguous, in general. When it is used for defining system specifications, this ambiguity may lead to misunderstandings between development teams that may result in increased costs and delays in development. If the misunderstandings are not detected, then a product that does not meet the intended specifications will be developed.

Ideally, specifications should be defined in a mathematical language, using formal logics. This not only removes ambiguity, but also allows system developers to utilize a vast set of methods [22] that have been developed by the academic community for testing and verification of systems. The academic community has also developed automatic tools such as S-TALiRO [2], [11], FAPAS [25], SpaceEx [9], CheckMate [19], FLOW [4], Breach [6], C2E2 [7], KeYmaera [18] and STRONG [5] that enable developers to conduct system testing and verification.

Even though it has been shown, that utilizing formal specifications can lead to improved testing and verification [8], the industry still utilizes natural language as the premier approach in defining specifications. One may conjecture that the most important reason for doing so is because the development of specifications through a formal logic requires a level of mathematical training that many users may not have [23]. Furthermore, even for expert users, writing formal specifications is an error prone task [10]. As a result, the industry has been less willing to utilize formal specifications in their processes.

In this work, we present a graphical formalism that enables non-expert users to develop formal specifications for control systems. The formalism enables the visualization of a large fragment of MTL. The main challenge in the development of the formalism lies in finding the right balance between expressive power and ease-of-use. It is designed for use with systems and signals and enables both event and time based specifications. This is the first time that a visual formal language representation is developed for specifications for Cyber-Physical Systems (CPS). Here by CPS we define any system that has discontinuous nonlinear dynamics and complex safety critical requirements. Prime examples are medical robotics and autonomous vehicles. A specification visualization tool has been developed based on the graphical formalism presented in this work. To evaluate the usefulness of the tool in terms of usability and ease-of-use, we have conducted a usability study.

SUMMARY OF CONTRIBUTIONS:

- We present a graphical formalism that enables the development of formal specifications.
- We present the visual specification tool based on the graphical formalism.
- We conducted a usability study to evaluate the tool.
- Through the usability study we proved that both non-expert users and expert users are able to define formal requirements accurately using the tool, and derived suggestions for improvement of the tool.
- We present applications of the tool for real-world robots.

RELATED WORKS: In order to help address the formal specification challenge, various graphical formalisms have been studied in the past [20], [1], [15], [3], [26], [21]. The most relevant works appear in [3] and [26]. In [3], the authors extend Message Sequence Charts and UML 2.0 Interaction Sequence Diagrams to propose a scenario based formalism called Property Sequence Chart (PSC). The formalism is mainly developed for specifications on concurrent systems. In [26], PSC is extended to Timed PSC which enables the addition of timing constructs to specifications.

In terms of usability studies for formal requirements very few works exist. In [23], the authors study the ability of expert users to develop requirements in Z. A related usability study for requirement representation is presented in [16], where the authors present and evaluate a system.
for generating, troubleshooting and executing controllers for robots using natural language.

II. VISUAL SPECIFICATION TOOL

The Visual Specification Tool (ViSPEC)\(^1\) enables the development of formal specifications for CPS. Users can develop requirements in a graphical formalism which is then translated to Metric Temporal Logic (MTL) [14]. The topic of capturing requirements through graphical formalisms has been studied in the past [20], [1], [15], [3], [26]. However, to the best of the authors knowledge, the work presented here is the first attempt to do so specifically aimed for the development of specifications for CPS. The initial idea for the graphical formalism was first presented in [11] while the tool was still in the early stages of development. However, in this work we present an updated version of the tool along with its usability study. The improvements over the previous version include: a more streamlined interface; an updated representation of signals in the interface; and an updated template definition process.

For CPS specifications, it is often needed to account for both timing and event sequence occurrence. Both of these are necessary for reasoning over systems and signals. Consider the specification \(\Box_{[0,5]}((\text{speed} > 100) \rightarrow \Box_{[0,5]}(\text{rpm} > 4000))\). It states that whenever within the first 5 seconds, the vehicle speed goes over 100, then from that moment on, the engine speed (rpm), for the next 5 seconds, should always be over 4000. Here both the sequence and timing of the events are of critical importance.

To ensure that the tool can be utilized by non-expert users, the following goals for the tool are defined: 1) The user interface is intuitive to use, i.e, it does not have a high learning curve; 2) The visual representation of the requirements is visually distinct and unambiguous; 3) There is a one-to-one mapping from the visual representation of the requirement and the corresponding requirement in MTL.

The set of specifications that can be generated from this graphical formalism is a proper subset of the set of MTL specifications. Formally, the following grammar produces the set of formulas that can be expressed by the proposed graphical formalism:

\[
\begin{align*}
S & \rightarrow \neg T \mid T \\
T & \rightarrow A \mid B \mid C \\
A & \rightarrow P \mid (P \land A) \mid (P \Rightarrow A) \\
B & \rightarrow \Box_I D \mid \Diamond_I D \\
C & \rightarrow \Box_I \Diamond_I D \mid \Diamond_I \Box_I D \\
D & \rightarrow p \mid (p \Rightarrow A) \mid (p \land A) \mid (p \Rightarrow B) \mid (p \land B) \\
P & \rightarrow p \mid \Box_I p \mid \Diamond_I p
\end{align*}
\]

where \(p\) is an atomic proposition. In practice, the atomic propositions are automatically derived from the templates.

Throughout the development process of the formalism, it was noticed that the more expressive the formalism, the more challenging to use it became. Therefore, we focused on several widely used classes of specifications which are described in Table I. Examples of the classes of specifications are presented in the rest of this section.

To make the tool easier to use, we placed several constraints on the types of signals used. Specifically, the signals and requirements are one dimensional. This enables clear and structured visualization on a two dimensional user interface.

In Fig. 1, the user interface of the tool is presented along with its most critical components. The user interface is composed of a menu, horizontal timeline, rectangular blocks called templates, and a zoom scroll. While the passage of time is represented horizontally, the sequence of events is presented vertically. The formulas are generated from templates as well as the connections between them.

The main building blocks of the formalism are templates. These are used for defining temporal logic operators, their timing intervals, and the expected signal shape. The user

\(^1\)Available at https://sites.google.com/a/asu.edu/s-taliro/vispec
starts with an empty template and a setup assistant presents
the user with a sequence of dialog boxes that aid in the
development of the template. The process is context dependent
where each option selection leads to a potentially different
set of options for the next step.

The first step in the template definition process is to
define the temporal operator. Among the choices (and their
 corresponding MTL symbols) are: Always (□), At Least
Once (◇), Eventually Always (◇□), Repeatedly Often and
Finally (∇), and now. The options available enable users to
define a wide range of specifications. The following sections
will present examples of a subset of formulas that can be
generated using this graphical formalism.

After the temporal operator is selected, the user sets
the timing bounds for it. Many users might have difficulty
defining timing bounds, especially for specifications with
temporal operators such as Eventually Always (◇□) and
Repeatedly Often and Finally (∇). To illustrate the process,
the tool provides a fill-in-the-blanks sentence format to
the user. For example, if the operator Eventually Always
is selected, the user will have to complete the following
sentence with the timing bounds: “ Eventually, between ____
and ____ seconds, the signal will become true, and from that
point on, will stay true in the next ____ to ____ seconds”.
The set of timing intervals are visualized with color shaded
regions in the template.

The next step in the process is in defining whether the
predicate will evaluate to true when the signal is above
or below a set threshold. For example, for the Always (□)
operator, a signal is selected that is either always above or
below a specified threshold. Once either option is selected,
various signals that fit the requirement are automatically
generated and presented visually. Instead of drawing the
signal, the user will select from one of the generated options.
Consider the following example:

**Example 1** A specification from the fragment of MTL
formulas called Safety MTL specifications is presented. Specif-
ically, the specification \( \phi_1 = □[0,36](\text{rpm} < 4000) \). The
formula states that in the next 36 seconds, engine speed
should always be less than 4000. The corresponding graph-
ical formalism for this formula is presented in Fig. 2. Note
that, in regards to the specification, the signal can be of any
shape as long as it is always below the 4000 threshold.

Consider the following example for the At Least Once (◇)
temporal operator:

**Example 2** A specification from the fragment of MTL for-
ulas called Reachability MTL specifications is presented.
Specifically, the specification \( \phi_2 = ◇[0,39](\text{speed} > 100) \).
The formula states that eventually, within the next 39 sec-
onds, the vehicle speed will go over 100. The corresponding
graphical formalism for this formula is presented in Fig. 3.
Again, in regards to the specification, the signal can be of
any shape as long as at one point, within the timing bounds
of the temporal operator, it is above the 100 threshold.

![Fig. 2: Example 1: The graphical formalism for the Safety
MTL specification \( \phi_1 = □[0,36](\text{rpm} < 4000) \).](image)

![Fig. 3: Example 2: The graphical formalism for the Reach-
ability MTL specification \( \phi_2 = ◇[0,39](\text{speed} > 100) \).](image)
Example 3 Consider the specification $\phi_3 = [0,30] [0,10] (speed > 100)$. The formula states that at some point in the first 30 seconds, the vehicle speed will go over 100 and stay above for 10 seconds. The corresponding graphical formalism for this formula is presented in Fig. 4.

For the Repeatedly Often and Finally ($\Box \Diamond$) operator, an oscillating signal is presented where two types of shading indicate the timing intervals for each MTL operator. Consider the following example:

Example 4 The specification $\phi_4 = [0,30] [0,10] (speed > 100)$ is presented. The formula states that at every timestep of the simulation in the first 30 seconds, the speed will go over 100 within the next 10 seconds. The corresponding graphical formalism for this formula is presented in Fig. 5. No matter how far to the left or right the green shaded region is moved, contained within the orange region, there is always a point where the signal is above the threshold. Recall that the signal is automatically generated so that it satisfies the options previously selected.

The next important concept in this graphical formalism is the relationship between templates.

First, the sequence relationship between two templates is presented. Assume that the first template is already created. If another template is added below it, then an order in the execution of the events is defined. The second template is only considered if the first template is evaluated to true. Formally, there is an implication relationship from the first template to the second. Consider the following example:

Example 5 The specification $\phi_5 = (\Diamond [0,40](speed > 100)) \rightarrow (\Diamond [0,30](rpm > 3000))$ is presented. The formula states that if, within 40 seconds, the vehicle speed is above 100 then within 30 seconds from time 0, the engine speed should be over 3000. The corresponding graphical formalism for this formula is presented in Fig. 6.

A second type of relationship enables the user to establish conjunction between two events. To achieve this, templates can be grouped. This is indicated by a bold black box. Doing so requires that both templates evaluate to true. Consider the following example:

Example 6 Specification $\phi_6 = ([0,40](speed < 100)) \land ([0,40](rpm < 4000))$. The formula states that, within 40 seconds, the vehicle speed should be less than 100 and the engine speed should be under 4000. The corresponding graphical formalism for this formula is presented in Fig. 7.

The third type of template relationship enables the user to establish relative timing between two templates. Consider the following example:

Example 7 Specification $\phi_7 = [0,40](speed < 80) \rightarrow [0,40](rpm < 4000)$). Here, the nested specification $[0,40](rpm < 4000)$ is evaluated every time $(speed < 80)$ is true. This formula is represented in the formalism with nested templates, otherwise referred to as parent and child templates. The second template is tabbed and connected to the first template using a green indicator. In the GUI, such a nested template is initiated by clicking on the signal of the parent template. The corresponding graphical formalism is presented in Fig. 8.

The variety of templates and the connections between them allow users to express a wide variety of specifications.
value threshold for the predicates. Secondly, the generated tree structure is traversed by a recursive algorithm to generate the MTL formula. There is a bijection between the visual representation of a specification and the MTL formula. An overview of the process is provided in Fig. 10.

An example of the tree structure for MTL formula \( \phi = \square(a \land \diamond b) \rightarrow (\square c \land \diamond(d \rightarrow (a \land \square b))) \) is shown in Fig. 11. The recursive algorithm for traversing the tree structure and generating the MTL formula is presented in Alg. 1. Note that the functions ADDPARENCOLL{A,B,C,D} add the parenthesis and connectives between predicates.

### IV. Usability Study

#### A. Hypotheses

The aim of the study is to evaluate whether ViSPEIC enables users to develop formal specifications. Two groups were considered:

1) Non-expert users: These are users who declared that they have no experience in working with requirements.
2) Expert users: These are users who declared that they have experience working with system requirements.

Some of the interesting questions we wanted to investigate, which are also presented as hypotheses in Tab. II, are:

- Whether the graphical formalism enables non-experts and experts to formalize requirements accurately.
- How well the expert cohort performs in comparison to the non-expert cohort.
- How user friendly and easy-to-use ViSPEIC is.

Writing formal requirements is a challenging task that requires a significant amount of training. Therefore, it is safe to assume that we can reject Hypothesis 1a as supported by our informal experience. Hypothesis 2a will be tested in a future work. In addition, we analyze user interaction and behavior to measure the ease-of-use of the tool.

**Algorithm 1** WriteMTL - Algorithm for generating the MTL formula given a tree structure of the graphical formalism

**Input:** Tree Structure \( T = (V,E) \) where \( v \in V \) and \( v = (G,Op,S) \) where \( G \) is the group, \( Op \) is the temporal operator and \( S \) is the predicate string; string \( \phi \)

**Output:** \( \phi \)

1: \[ function \text{ WRITEMTL}(T, \phi) \]
2: \[ C \leftarrow T.getChildren. \]
3: \[ sC \leftarrow \text{size}(C) \]
4: \[ for \ i \ in C \ do \]
5: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'i.Op}) \]
6: \[ if \ i.isParent \ then \]
7: \[ if \ not(i.S.isEmpty) \ then \]
8: \[ subC \leftarrow \text{t.getChildren}(i) \]
9: \[ if \ i.G == subC(1).G \ then \]
10: \[ \phi \leftarrow \text{CONC}(\phi, \text{'','i.S,'\&'}') \]
11: \[ else \]
12: \[ if \ i.isParent \ then \]
13: \[ \phi \leftarrow \text{CONC}(\phi, \text{'','i.S,','\&'}') \]
14: \[ else \]
15: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'i.S,'->'}) \]
16: \[ end if \]
17: \[ end if \]
18: \[ \phi \leftarrow \text{WRITEMTL}(i.subtree, \phi) \]
19: \[ if \ i.isParent \ then \]
20: \[ if \ i.G == subC.G \ then \]
21: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
22: \[ else \]
23: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
24: \[ end if \]
25: \[ else \]
26: \[ if \ sC > 1 \ and \ i \neq sC \ then \]
27: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'->'}) \]
28: \[ else \]
29: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
30: \[ end if \]
31: \[ else \]
32: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
33: \[ \phi \leftarrow \text{WRITEMTL}(i.subtree, \phi) \]
34: \[ if \ i \neq sC \ then \]
35: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'->'}) \]
36: \[ else \]
37: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
38: \[ end if \]
39: \[ end if \]
40: \[ else \]
41: \[ \phi \leftarrow \text{CONC}(\phi, \text{,''}) \]
42: \[ \phi \leftarrow \text{WRITEMTL}(i.subtree, \phi) \]
43: \[ if \ i \neq sC \ then \]
44: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'&'}) \]
45: \[ else \]
46: \[ \phi \leftarrow \text{CONC}(\phi, \text{,'->'}) \]
47: \[ end if \]
48: \[ end if \]
49: \[ end for \]
50: \[ end function \]

![Fig. 7: Example 6: The graphical formalism for the MTL specification \( \phi_6 = (\square[0,40](speed < 100)) \land (\square[0,40](rpm < 4000)) \).](image-url)
TABLE II: Hypotheses and test results with level of significance $\alpha = 0.05$. User groups as defined in section IV.A.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Reject null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a Non-expert users are able to define formal requirements accurately using formal logics such as MTL.</td>
<td>Yes</td>
</tr>
<tr>
<td>1b Non-expert users are able to define formal requirements accurately using the Visual Specification Tool.</td>
<td>Yes</td>
</tr>
<tr>
<td>2a Expert users from the industry are able to define formal requirements accurately using formal logics such as MTL.</td>
<td>Yes</td>
</tr>
<tr>
<td>2b Expert users from the industry are able to define formal requirements accurately using the Visual Specification Tool.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

$T_{x_{alt}}$ The mean grade per task $x$ for industry users is greater than to the mean grade per task $x$ for non-expert users. Partially

---

Fig. 8: Example 7: The graphical formalism for the MTL

$\phi_7 = \square_{[0,40]}(\text{speed} < 80) \rightarrow \square_{[0,40]}(\text{rpm} < 4000)$.  

B. Demographics

The non-expert cohort was comprised of twenty subjects from the student community of Arizona State University. Most of the subjects are from an engineering background with little to no experience working with requirements. The student demographics are presented in Tab. III.

The subject cohort comprised of ten subjects from the industry in the Phoenix area. The subjects have experience working with specifications and come from an engineering background.

TABLE III: Hypothesis $1b$ Subject Demographics

| Freshman | 2 | Computer Science | 5 | Male | 12 |
| Sophomore | 2 | Software Engineering | 3 | Female | 8 |
| Junior | 5 | Electrical Engineering | 3 |       |    |
| Senior | 5 | Mechanical Engineering | 6 |       |    |
| Masters | 4 | Engineering, other | 3 |       |    |
| PhD | 2 |                        |    |       |    |

C. Experimental Design

Each subject received a task list to complete. The task list contained ten tasks related to automotive system specifications. Each task asked the subject to formalize a natural language specification through ViSPEC and generate an MTL formula. The list of tasks is presented in Table VI.

The tasks become more complex throughout the session. The higher the number of the task, the more steps necessary to complete the task successfully.

Each session is at most 45 minutes long. Subjects received a one minute and thirty second tutorial on using ViSPEC to develop specifications. The computer screen was recorded and actions were logged for each session. The subjects also completed a demographic and post-completion questionnaire.

D. Metrics

Two metrics are used for performance evaluation:

Task completion: this is a binary measure, which indicates whether users were able to finish the task within the set time.

Measure of Accuracy: a value from one to five which is used to quantify the accuracy of subject generated formulas. The formulas are graded by formal specification experts which were given the following two suggested criteria: a) How accurate the meaning of the natural language specification is captured, and b) Whether the inaccuracies in the user submitted formula can be easily debugged and corrected in the testing and verification process. Furthermore, in order to decrease subjectivity, the following instructions were given to the expert graders in order to anchor the meanings of the five different grades of the scale used: A grade of one indicates that the generated formula is totally inaccurate. A grade of two indicates that the formula is mostly inaccurate. A grade of three indicates an inaccurate formula which can be easily debugged and corrected to the proper formal logic specification by formal specification experts and thus this is the minimum acceptable satisfactory result. A grade of four indicates that the formula is inaccurate but can be debugged and improved by automated specification debugging tools. A grade of five indicates that the generated formula is completely accurate. The group of expert graders consisted of experts in formal methods and logic.

V. RESULTS

1) Average grade per task: For both cohorts, the task performance is presented in Fig. 9. It can be observed that overall, the mean grade per task for both cohorts is high. Consider the mean grade per task as a random variable $\bar{X}$. Specifically, $\bar{X} : \Omega \rightarrow \mathbb{R}$, where $\Omega \in \{ y : 1 \leq y \leq 5 \}$. In Figure 12, we present the survival function $S_{\bar{X}}(x) = 1 - F_{\bar{X}}(x) = 1 - P(\bar{X} \leq x) = P(\bar{X} > x)$ based on sample data. Note that $x$ is the threshold of mean grade accuracy.

2) Hypothesis 1b: To test Hypothesis $1b$, we need to establish what is an acceptable threshold for accuracy in order to test the hypothesis. As discussed in the metrics section, we claim that a mean grade higher than three is an acceptable threshold for non-expert users. Therefore, hypothesis $1b$ is reduced to the null hypothesis: the mean grade per user is less than or equal to three for non-experts.

Let us define the average grade per user as a random variable $\bar{Y}$. Specifically, $\bar{Y} : \Omega \rightarrow \mathbb{R}$, where $\Omega \in \{ y : 0 \leq y \leq 5 \}$.
5) **Hypothesis T**: We test Hypothesis T for the non-expert and expert users. We conduct a two sample t-test. The p-value returned from the test is 0.0024 and for a significance level of 0.01, we reject the null hypothesis. Therefore we claim that the mean grade per user for expert users is greater than the mean grade per user for non-experts.

We can use the t-statistic to test the hypothesis. We reject the null hypothesis with a p-value very close to 0. 

Next, we compare the mean grade of three tasks for expert users. We test for normality as in the previous case and all three test fail to reject the null hypothesis that the data follows the normal distribution. In figure 14, we plot the non-expert data against a normal distribution and the corresponding Q-Q plot. If we assume that the data constitute a random sample from a normal distribution, i.e., $Y \sim \mathcal{N}$, we can use the t-statistic to test the hypothesis. We reject the null hypothesis with a p-value very close to 0.

Similarly, we test Hypothesis 2b for the expert cohort. Hypothesis 2b is reduced to the null hypothesis: the mean grade per user is less than or equal to three for expert users. We test for normality as in the previous case and all three test fail to reject the null hypothesis that the data follows the normal distribution.

Consider the average grade per user as a random variable $\bar{Z}$. Specifically, $\bar{Z} : \Omega \rightarrow \mathbb{R}$, where $\Omega \in \{y : 1 \leq y \leq 5\}$. The sample data from 10 subjects has a mean grade of 4.76 and standard deviation of 0.26. In figure 14, we plot the non-expert data against a fitted normal distribution and the corresponding Q-Q plot. If we assume that the data constitute a random sample from a normal distribution, i.e., $\bar{Z} \sim \mathcal{N}$ we can use the t-statistic to test the hypothesis. We reject the null hypothesis with a p-value very close to 0.
TABLE IV: ViSPEC improvements

<table>
<thead>
<tr>
<th>#</th>
<th>Improve...</th>
<th>Prime Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>the process of creating child templates</td>
<td>misclicks; user feedback</td>
</tr>
<tr>
<td>2</td>
<td>the tutorial by placing more emphasis on the difference between implication and conjunction between templates</td>
<td>task accuracy grade</td>
</tr>
<tr>
<td>3</td>
<td>the visual representation of grouped templates</td>
<td>task accuracy grade; user feedback</td>
</tr>
</tbody>
</table>

presented in Tab. V. Task 9 is the most difficult task when it comes to the number of errors generated, and this is the only task where there is a clear difference in performance between the expert and non-expert cohorts.

TABLE V: Hypothesis testing of $Tx_{null}$ with $\alpha = 0.05$

<table>
<thead>
<tr>
<th>$x$</th>
<th>Rej. $Tx_{null}$</th>
<th>p-val.</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>No</td>
<td>0.065</td>
<td>potentially true with more investigation</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>0.165</td>
<td>false</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>0.074</td>
<td>potentially true with more investigation</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>0.100</td>
<td>potentially true with more investigation</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>0.424</td>
<td>false</td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td>0.016</td>
<td>true</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>0.063</td>
<td>potentially true with more investigation</td>
</tr>
</tbody>
</table>

We observe that the only null hypothesis rejected is for task nine indicating that the mean grade for expert users is greater than the mean grade for non-expert users. The subject accuracy grades over tasks for is shown in Fig. 9.

6) Ease-of-use analysis: One indicator for the ease-of-use of the application is the total time spent per task. As can be observed in Fig. 13, the mean time spent per task on average is at most 167 seconds. For easier identification of points of difficulty, we divided each task into subtasks. It was observed that there is no correlation between the length of time spent in a subtask and correctness. This potentially indicates, as also verified by correlation testing between times and grades, that the subjects were unaware of mistakes in the process. From these and other observations, such as misclicks, and subject feedback, we have developed a set of refinements on the tool to improve the user experience. A partial list of improvements is presented in Table IV.

VI. APPLICATIONS

A. Robotic Surgery

In the last few decades, there has been a significant increase in the number of robotics systems, especially in the health care system. They have been successfully introduced in multiple areas such as rehabilitation, telesurgery, physical therapy, elderly care, and remote physician care. In the following, we will focus on autonomous robotic systems for surgery where of paramount importance is the safety of these systems [13]. Specifically, we will consider a model of a robotic serial link manipulator as presented in [17].

One of the main tasks in surgery is the puncturing action. The high precision and repeatability of the process, make robot systems ideal for this task. Also, the trauma induced around the region is much lower and therefore the recovery process for the patient is quicker. To complete the puncturing action, the robot has to move towards the puncturing location.

Test the tissue for various indicators to calibrate for optimal puncture, bring the puncturing needle to a perpendicular position and, finally, puncture with correct force and angle. If the force or angle is miscalculated, it might pose unintended harm to the patient. Consider the specifications from [17] that should hold on a serial manipulator for puncturing:

1) From [17]: The force applied to the patient by the end effector is always less than a given threshold, except for the puncturing subtask. Formally, assuming that the operation time is 30 seconds, we have: $\phi_{s1} = \circ_{[0,30]}(\neg\text{puncturing} \rightarrow f \leq f_{\text{max}})$.

2) From [17]: The task is feasible, and the position of the needle once it stops is inside the target region $R$. Formally, assuming that the operation time is 40 seconds, we have: $\phi_{s2} = \diamond_{[0,40]}(\text{Stop} \land \text{needle} \in R)$.

3) Also, other requirements can be expressed for such a system. For example, the end effector speed should not be less than $v_{\text{min}}$ and should not be greater than $v_{\text{max}}$. Formally: $\phi_{s3} = \circ_{[0,40]}(v_{\text{min}} < v_{\text{ef}} < v_{\text{max}})$.

The ViSPEC tool is utilized to develop the specifications for the robotic manipulator. For $\phi_{s1}$, the specification is presented in Fig. 15. We assume that $f_{\text{max}} = 10$. For $\phi_{s2}$, the specification is presented in Fig. 17. We assume that $\text{needle} \in R \Leftrightarrow 5 < n_x < 10 \land 5 < n_y < 10$, where $n_x, n_y$ are the $x$ and $y$ coordinates for the needle. For $\phi_{s3}$, the specification is presented in Fig. 18. We assume that $v_{\text{min}} = 10$ and $v_{\text{max}} = 20$.

B. Quadcopter

In recent years, quadcopters and other unmanned aerial vehicles (UAVs) have become a major focus for research both in the academic community and industry. Among others, they are used in military operations, nuclear disaster assessment, firefighting and entertainment. The challenges faced in developing these devices and their control algorithms come from the flight dynamics and the highly dynamical environment that they operate in. Also, as the complexity of these devices increases, so do the performance and reliability requirements.

Consider the following specifications for a quadrotor:

1) The absolute value of the pitch and roll angle should always be below certain thresholds. Formally, assuming that the operation time is 40 seconds, we have: $\phi_{q1} = \circ_{[0,40]}(|\alpha| < \alpha_{\text{max}}) \land \circ_{[0,40]}(|\beta| < \beta_{\text{max}})$.

2) If distance to the target region is smaller than a certain threshold $d$, then for the next 20 seconds, the speed should not exceed $v_{\text{max}}$. Formally, assuming that the operation time is 40 seconds, we have: $\phi_{q2} = \circ_{[0,40]}(\text{dist} < d \rightarrow \circ_{[0,20]}(v < v_{\text{max}}))$.

The ViSPEC tool is utilized to develop the specifications for the quadrotor. For $\phi_{q1}$, the specification is presented in Fig. 16. We assume that $\alpha_{\text{max}} = 45$ deg, $\beta_{\text{max}} = 45$ deg and $\gamma_{\text{max}} = 60$ deg. For $\phi_{q2}$, the specification is presented in Fig. 19. We assume that $d = 5$ and $v_{\text{max}} = 10$. For $\phi_{s3}$, the specification is presented in Fig. 18. We assume that $v_{\text{min}} = 10$ and $v_{\text{max}} = 20$. 
TABLE VI: Task list with automotive system specifications presented in natural language

<table>
<thead>
<tr>
<th>Task</th>
<th>Natural Language Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Safety</td>
<td>In the first 40 seconds, vehicle speed should always be less than 160.</td>
</tr>
<tr>
<td>2. Reachability</td>
<td>In the first 30 seconds, vehicle speed should go over 120.</td>
</tr>
<tr>
<td>3. Stabilization</td>
<td>At some point in time in the first 30 seconds, vehicle speed will go over 100 and stay above for 20 seconds.</td>
</tr>
<tr>
<td>4. Recurrence</td>
<td>At every point in time in the first 40 seconds, vehicle speed will go over 100 in the next 10 seconds.</td>
</tr>
<tr>
<td>5. Recurrence</td>
<td>It is not the case that, for up to 40 seconds, the vehicle speed will go over 100 in every 10 second period.</td>
</tr>
<tr>
<td>6. Implication</td>
<td>If, within 40 seconds, vehicle speed is above 100 then within 30 seconds from time 0, engine speed should be over 3000.</td>
</tr>
<tr>
<td>7. Reactive Response</td>
<td>If, at some point in time in the first 40 seconds, vehicle speed goes over 80 then from that point on, for the next 30 seconds, engine speed should be over 4000.</td>
</tr>
<tr>
<td>8. Conjunction</td>
<td>In the first 40 seconds, vehicle speed should be less than 100 and engine speed should be under 4000.</td>
</tr>
<tr>
<td>9. Non-strict sequencing</td>
<td>At some point in time in the first 40 seconds, vehicle speed should go over 80 and then from that point on, for the next 30 seconds, engine speed should be over 4000.</td>
</tr>
<tr>
<td>10. Long sequence</td>
<td>If, at some point in time in the first 40 seconds, vehicle speed goes over 80 then from that point on, if within the next 20 seconds the engine speed goes over 4000, then, for the next 30 seconds, the vehicle speed should be over 100.</td>
</tr>
</tbody>
</table>

![Non-expert data fit with Normal distribution](image1)

![Expert data fit with Normal distribution](image2)

**Fig. 14:** Subject data fit with a normal distribution and the corresponding Q-Q plot.

![QQ Plot of Sample Data versus Standard Normal](image3)

**Fig. 15:** The graphical formalism for $\phi_{s1}$.

**VII. CONCLUSION AND FUTURE WORK**

As robots and other cyber-physical systems become more complex and ubiquitous, so does the need for better testing and verification. A set of formal methods that improve this process require some formal representation of system specifications. In this work, a graphical formalism and a tool that enables users to easily develop formal specifications are presented. The ViSPEC tool enables users who have little to no mathematical training in formal logics to develop formal specifications, as was verified by a usability study that was conducted in order to evaluate the usefulness of the tool and to get insights on potential improvements. The tool was utilized to formalize specifications for two robots.

Last but not least, we would like to investigate if the potential inaccuracies of the specifications that users generate with the tool can be attributed mainly to the inherent ambiguity of the natural language descriptions which were given, or if not, which other factors contribute and to what extent. Thus, in an improved usability study, we aim towards exploring alternative methods of generation of requirements from engineers for a system, that do not involve the administration of a natural language description by the experimenter. This would enable us to study to what extent inherent natural language ambiguity causes the observed less-than-perfect accuracy that is sometimes, even if rarely, exhibited.

**ACKNOWLEDGMENT:** Partial support under NSF awards CNS-1319560, CNS-1116136, IIP-1454143, IIP-1361926 and the NSF I/UCRC Center for Embedded Systems. We thank all the participants in the usability study and the
reviewers for the detailed reviews.

REFERENCES


