Narrative Monologue as a First Step Towards Advanced Mission Debrief for AUV Operator Situational Awareness

Nicholas A. R. Johnson and David M. Lane

***Abstract*— Present generation AUVs act as mobile sensor platforms, allowing post mission debriefs which require little more than an adequate expression of the collected data, and only basic requirements for mission planning. As the need and capability advances, autonomy will increase in turn, leading to more complex missions, which must be supported by more capable operator interfaces, both for post mission debrief and pre-mission verification, lest a deficit of trust form between the vehicles and their operators.**

We present *Glaykos*, a system designed to automatically create audio visual debriefs for missions carried out with both real and simulated vehicles, allowing it to act as a tool for both pre-mission verification and post mission debrief. A series of data transforms are applied to the domain information and mission logs from an AUV mission, resulting in the creation of an audio visual debrief whose structure is informed by principles of human computer interaction and narrative. The approach is tested for effectiveness and results of this are presented, showing that it better able to keep users informed when compared two other debriefing strategies, one of which mimics the current state of the art. Conclusions are drawn regarding the system's monologue based approach acting as a first step towards more complex and capable multimodal, dialogue based operator interfaces for autonomous systems.

I. Introduction

AUVs have found a place in the military toolkit and have been successfully deployed in various scenarios, particularly those related to Mine Counter Measures[1]. These same strengths are now leading to the adoption of AUVs for complex tasks in the offshore oil industry, providing more efficient methods of carrying out tasks which would otherwise require ship borne Remotely Operated Vehicles (ROVs) or towed side scan sonar platforms (or "tow fishes")[2], [3]. Work in academic laboratories, such as the Ocean Systems Laboratory at Heriot-Watt University, continues to advance the field further, increasing the scope and capabilities of AUVSs[4], [5], [6], [7], [8].

Communication in the case of AUVs has a high latency and an extremely low bandwidth. As a result, the only opportunities for high bandwidth data exchange between the operator and the AUV are before and after the mission. Additionally, the operator must make do with a bare minimum of status updates during the mission, which may contain little more than the position of the vehicle. As a result, a higher degree of intelligence must be placed on the vehicle and the operator must trust this to be able to correctly carry out the mission. As more complex requirements emerge for AUVs, so they develop a requirement for a larger degree of autonomy, and the issue of trust becomes even greater still.

Current generation commercial solutions for user interaction with AUVs[9], [10], [11] reflect the fact that most current generation AUVs are essentially deployed as sensor platforms, with missions consisting of a static list of waypoints, and are designed with this application in mind. Missions are planned in their entirety by the operator on an entirely spatial basis. These systems have little scope or integrated functionality for the verification of the mission after it has been planned. Likewise, the post mission interaction is largely limited to the path taken by the vehicle and sensor data collected whilst moving along it.

As the state of the art for autonomous systems advances, however, so will the need for more advanced user interaction and multimodal interfaces to supply this human robot interaction. Coupled with this will be requirements for more effective pre-mission verification and post mission debrief, as crucial tools which will be needed to foster trust in operators for AUV systems[12], [13], [14]. Some of our previous work[15] described a framework for the creation of such an interface system.

A key tool for the crucial output processing component of this framework would appear to be narrative. One study[16] attempted to identify the effect of numerous variables on their subject's ability to recall prose passages. Retention scores for narrative texts were found to be significantly higher than those for expository texts. Subsequent work by Pennington and Hastie[17], [18], [19] discovered and demonstrated a different benefit of narrative, in an entirely different arena. In this work, they develop and test the "Story Model", an explanation based theory for juror decisions, and then test this by gauging the effects of various different factors on the decisions made by jurors. As a result they found, as they had posited, that the factor which had the largest effect on jurors' decisions, and their confidence in these decisions, was the coherence of the narrative which was presented to them.

Presented herein is a system designed to act as the first step towards more powerful operator interfaces for autonomous systems. It automatically generates an audio visual (including synthesized vocal narration) debrief for complex missions carried out by autonomous assets and plays this to the user in the form of a narratively structured monologue. This monologue is designed to act as the first step to towards an eventual advanced debriefing system capable of participating in a dialogue with the user.

This work was carried out with the support of the Engineering and Physical Sciences Research Council, and SeeByte Ltd.

Nicholas A. R. Johnson was with the Ocean Systems Laboratory at Heriot-Watt University, Edinburgh (and is now with Google UK Ltd.) narj1@hw.ac.uk,nick.johnson1981@gmail.com

David M. Lane is with the Ocean Systems Laboratory at Heriot-Watt University, Edinburgh d.m.lane@hw.ac.uk

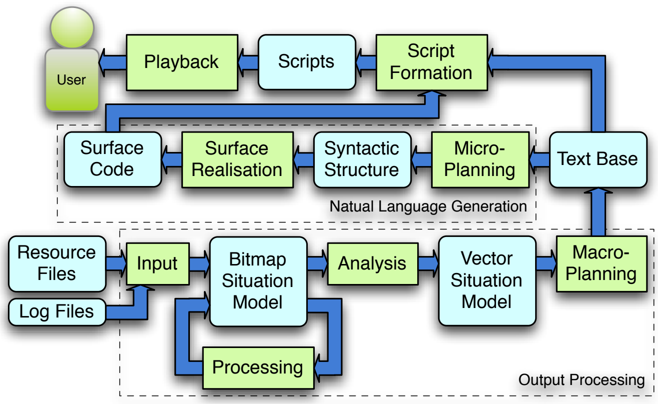

Fig. 1. Data Flow in the *Glaykos* System. The sections outlined with dashed lines refer back to the framework described in [15]. The system forms a pipeline of data transformations, starting with the raw domain information and mission logs, and ending with a complex audio visual debrief (including synthesized verbal narration). Two levels of situation model are used to abstract the data, before it is ordered and then realised for delivery to the user. This process follows the levels of representation outlined in [21].

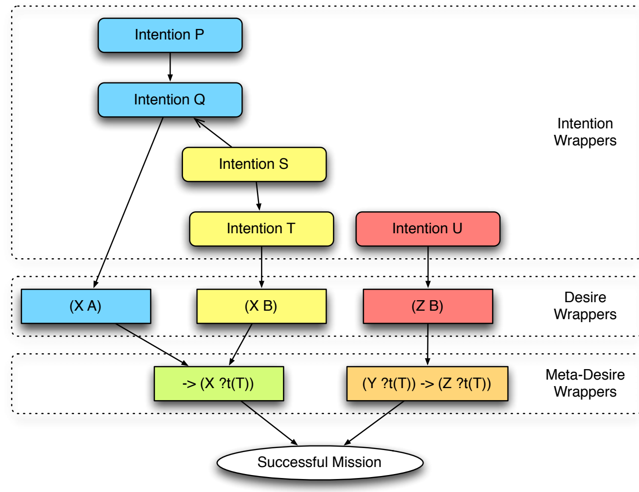

Fig. 2. Example motivation model, the first of the acyclic, directed graphs which make up the vector situation model. The arrows represent the dependencies between the individual components.

II. Implementation

In the currently implemented system, a prototype deliberative control system based on the FF planning system[20] is used with simulated vehicles. Simulation is run at ultra high speed in order to enable rabid iteration of the mission parameters as part of a pre-validation procedure.

The Glaykos system takes the form of a pipeline of data transformations, starting with the provided domain specificiation and logs output from the vehicles post mission, and ending with a complete audio visual debrief. This pipeline is summarised in Figure 1.

In order to create the debrief, the levels of representation stated in [21] are followed. First a complete situation model is built in the computer. This is done in two steps. The first is the **bitmap situation model**, which contains all of the data from the mission, and consists of a large and highly connected network of simple concepts. In the current implementation is the model is stored in an ontology, as this allows the easy creation of complex and powerful searches over the knowledge base. The data from the domain model and mission logs is fed into this, and then processed further in order to develop all of the connections between each element of the model.

From this is built the **vector situation model**, which is made up of a much smaller number of more complex concepts. These are contained within two acyclic directed graphs, which model the motivation and causation of the mission. The first and simplest of these graphs is the motivation model, which describes the relationships between each of the mission goals and the intentions which are added to the plan in order to achieve them. Figure 2 gives a simple example motivation model.

The second of the graphs in the vector situation model is the causation model, and is made up of **process elements**, which describe events which occur at a particular moment in time. These are a symbolic representation of a "process statements", one of the two fundamental building blocks of narrative described in *Story and Discourse*[22]. An example process statement might be: "The Manipulator AUV arrived at the base location." Each of these are positioned on the five situational axes: space, time, protagonist, motivation and causation, as described in [21] and[23]. This is generated from the data contained in the motivation model and further details taken from the bitmap situation model. It described the causal connections between each event in the mission. Figure 3 provides a simple example causation model, the colouring of which matches the example motivation model.

The next stage of the process is to arrange the process elements which make up the causation model into a concrete ordering for delivery to the use. In order to ease this process it is first simplified by grouping elements together where appropriate. Several grouping strategies are employed, but in general any elements which have the following relationship are grouped together:

- They are chronologically adjacent ;
- They all have the same protagonist;
- They do not have differing primary motivations;
- The first element is the cause of the second *and no others*.

This grouping is shown in Figure 3. These groups are then arranged into a concrete ordering by optimisation algorithms which treat the task as an instance of the asymmetric travelling salesman problem. Two algorithms are deployed for this purpose, a relatively simple bounded and constrained depth first search, and a random key based genetic algorithm[24]. In each case, the fitness function is the same, and is based on giving quantitative values to the penalties described in [21] for discontinuity on the five situational axis:

1) **Spatiality**. The action or event expressed in the Statement is in a different spatial region from the content

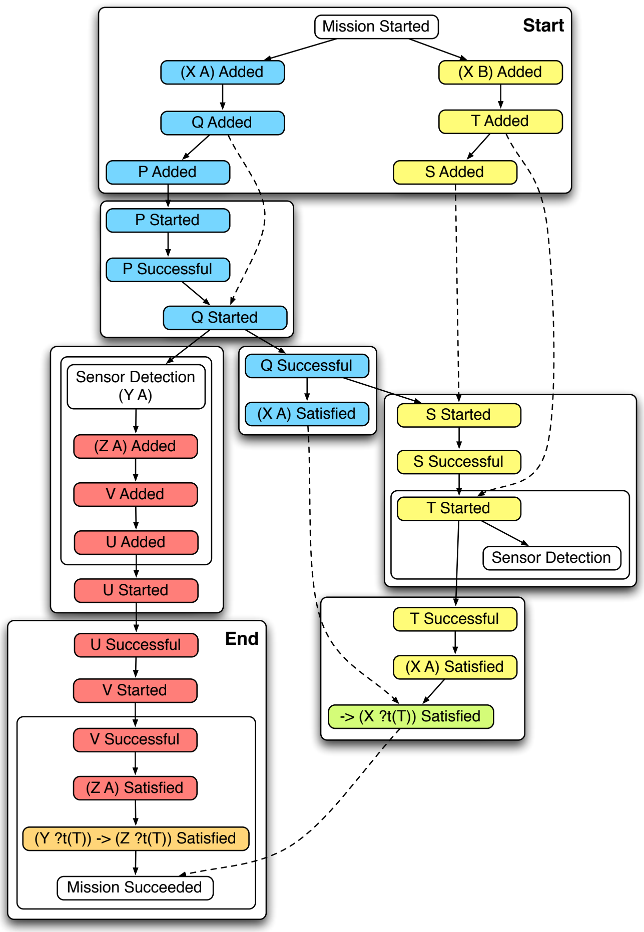

Fig. 3. Grouped example causation model, the second of the acyclic, directed graphs which make up the vector situation model. The smaller boxes are the process elements which represent the events of the mission. The colours and text refer back to the motivation model given in Figure 2.

in Working Memory.

2) **Temporality**. The action or event expressed in the Statement involves a gap or shift in the chronological timeline.
3) **Protagonist**. The protagonists in the Statement are not among the protagonists in Working Memory.
4) **Motivation**. The action expressed in the Statement is not part of an agent's plan in Working Memory.
5) **Causality**. The action or event expressed in the Statement does not causally flow from the content in Working Memory.

This ordering forms the basis for the **text base**, which is then annotated with **stasis elements** as required. These are symbolic representations of the second fundamental building block of narrative (the "stasis statement") from [22], and provide additional contextually relevant expository information. Stasis statements describe information which is true for a period of time. An example might be: "The Repair AUV has the repair installations module." The ordered list of process and stasis elements forms the complete text base.

Each of these elements is translated to natural language using a natural language generation system based on SimpleNLG[25]. Each element of the domain model is annotated with the additional information information required to do this. The elements which make up each of the beliefs, desires and intentions used by the planning system can them be converted into a symbolic form. The template syntax tree which is used for the desires is shown in Figure 4. These can then be used as parts of the syntax trees which describe the process and stasis elements, such as the desire satisfied process element, the syntax tree for which is shown in Figure 5.

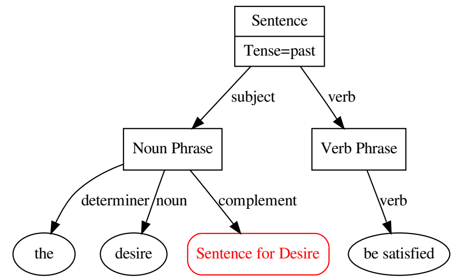

Fig. 5. Syntactic structure for "Desire Satisfied" process statement.

These syntax trees are then converted to natural language, which yields the **surface text**, in the form of an ordered set of sentences. Spoken language is generated using a speech synthesizer, and the time taken for each element recorded. This is used to build the the debrief's model of time during the replay.

For each process statement, relevant graphic representation of portions of the motivation model are also generated to be shown to the user as part of a heads-up-display used during the audio visual debrief. Depending on the nature of the process statement being related to the user, these can take two forms.

If the current statement relates directly to the mission goals, a desire based representation is used (as per the example in Figure 6), which shows the mission's goals tree and the status of the goals within it.

If the statement relates the actions which are being carried out by the AUVs, then the intention based representation is used (as per the example in Figure 7) which shows the section of the plan which is working towards the current goal.

All transitions between motivation graphics are animated in order to ensure that the connection between each graphic is made as clear as possible. Finally, Figure 8 shows the additional elements which are displayed to the user on the heads-up-display during the mission debrief. The debrief is delivered as a monologue, with each statement being spoken by the speech synthesiser (with subtitles) while the system controls the timing of the mission replay and displays the generated motivation graphics. In the current system, no attempt is made to control the "camera" and move the user's spatial viewpoint around the mission area.

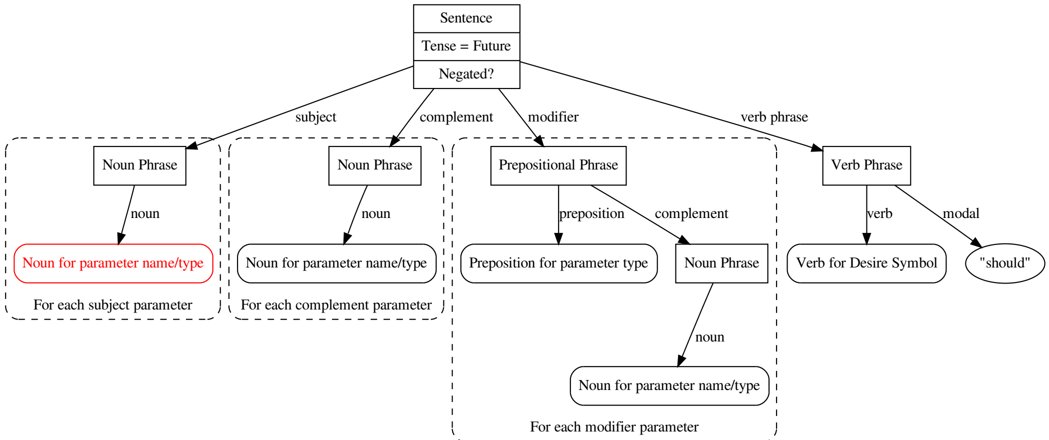

Fig. 4. Syntactic structure for desires. Structures in red are those which may be converted to pronouns by the natural language generation system.

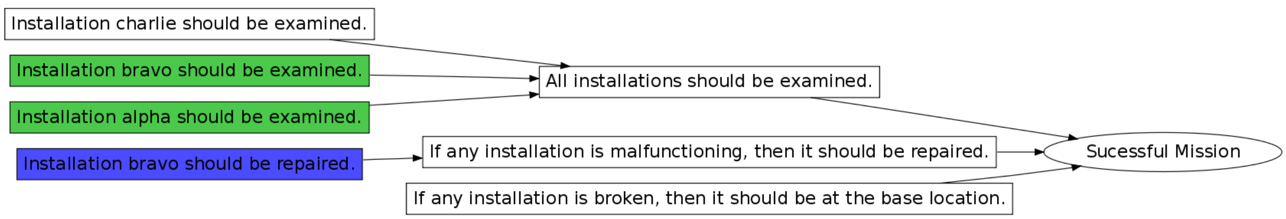

Fig. 6. Example of a graphic used to display the current state of the mission goals to the user.

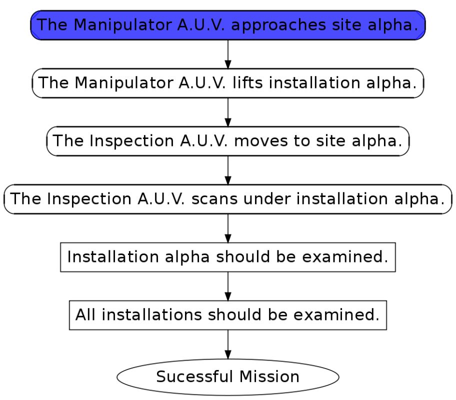

Fig. 7. Example of graphic used to display an AUVs current plan to the user.

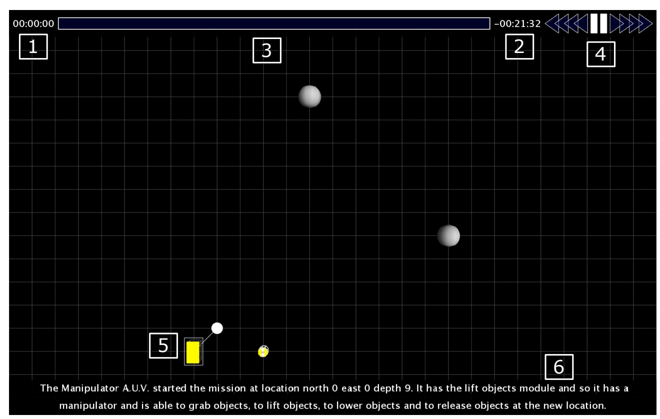

Fig. 8. Example of features in the Heads Up Display. Including: 1) Mission time passed; 2) Mission time remaining; 3) Mission timeline; 4) Mission playback speed indicator; 5) Currently selected agent; and 6) Subtitle.

III. Testing

Three scenarios were used in the testing of the Glaykos system. An offshore maintenance based task in which vehicles were required to directly co-operate in order to examine and repair several offshore installations. A simple search/classify/neutralize mine counter measures task, and a combination scenario which combined the goals of the first two missions.

Debriefs were provided to participants using one of three experimental cases:

- The debrief generated by the Glaykos system (experimental case);
- A similar, but chronologically ordered, debrief, with no intention graphics (base case);
- A system similar to the current state of the art, which gives complete control over spacial and temporal playback, but no guidance (user case).

Participants were divided into six groups, with the first two viewing the offshore and MCM scenarios with either the base or experimental case (both groups experiencing bother scenarios and case). Likewise, the third and fourth groups used the experimental and user cases. The final two groups (which consisted of a smaller number of domain experts)

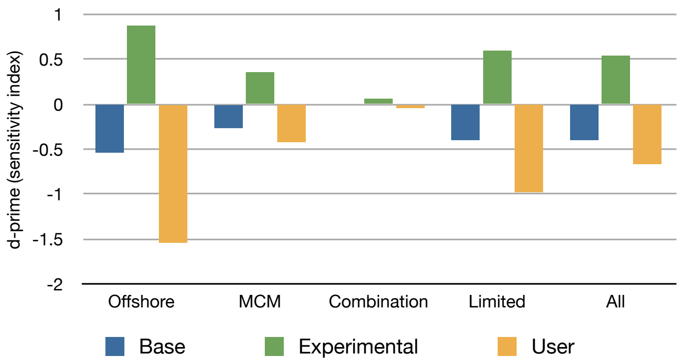

Fig. 9. The average *d-prime* figures for each scenario and case, without the inclusion of the data from Participant 13. "All" averages all scenarios for each case, while "limited" excludes the combination scenario, for which a limited number of data points were available.

viewed the combination scenario with either the experimental or user case.

Post debrief surveys were then used to assess each participant's subjective and objective reactions to the debriefs. The user case also collected a complete record of the participants' interaction with the debrief. Between the debrief and the survey, each participant viewed a short animated film.

IV. Results

One of the questions on the post debrief surveys asked the participant to repeat the events of the mission they had viewed, in whichever order they thought was most appropriate. These data was then compared to important facts about each scenario in order to asses each participant's rate of hits, misses and false alarms. This was then used to calculate the *d-prime* (or sensitivity index) measure, which can be used to asses the quality of the recall of events. The average *d-prime* value for each of the scenarios[1], together with combined values can be found in Figure 9.

It can be seen from this graph that the average *d-prime* is higher for the *experimental case* than either the *base* or *user case* in every instance (though by only a small amount in the case of the combination scenario). The individual d-prime figures were tested for statistical significance using Student's t-test, the results of which are shown in Table I. This shows that participants accuracy was significantly higher for the *experimental case* with the offshore maintenance scenario, and both combined measures, though not for the MCM or combination scenarios taken in isolation.

One of the key differences between the *base* and *experimental cases* is the fact that the *experimental case* eschews chronology in favour of an ordering more closely tied to the causal relationships between the events. This proved to be quite contentious, with the majority of participants making comment upon it in the survey, as shown in Figure 10. The largest group of these were in the negative camp, though a

[1]One participant has been removed from these data, as they revealed after the completion of the experiment that they had come directly from a long haul flight and were later found to represent an extreme outlier.

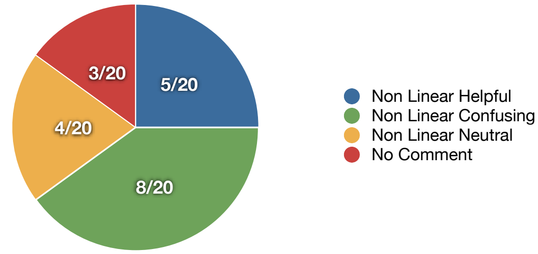

Fig. 10. Participant opinion of non-linear ordering.

reasonable portion either specifically stated that they found this ordering helpful or were ambivalent regarding the matter.

Interestingly, though, the majority of participants chose not to use a chronological ordering when relating the the events of the mission as part of the previously described question. Three distinct orderings were used by the participants:

- **Chronological**, in which the events are listed in a strict chronological ordering.
- **Causal**, in which the events are given an ordering which tends to prioritise causal connections over temporal ones.
- **Motivational**, which takes a higher level view, and orders the events according to the particular goal they satisfy.

Figure 11 gives the distribution of orderings used by the participants when describing the offshore and MCM scenarios. In general there was a large amount of variation present in the orderings the participants chose to use, with chronological proving to be the least popular overall and the two non-chronological orderings proving equally popular. Something which is implicit in these graphs is that participants who viewed two scenarios did not necessarily use the same ordering when describing each. This suggests that the best ordering of events is a function of both the individual participant and the scenario itself.

V. Conclusion

The results given here show the the debrief generated by the Glaykos system was better able to keep the user informed than either a simpler chronological debrief or a system emulating the current state of the art. Though a large portion of participants found the non-chronological ordering which was employed to be confusing, others found it actively helpful. Interestingly, the numbers show that at least some of those who said they found a non-chronological ordering to be confusing actually used a non-chronological ordering when asked to recount the events of the mission in question.

When considering the results, it is important to note, first of all, that the system described here was never meant to represent any kind of complete solution, or finished product. The current system is intended to act as a functional prototype, and a stepping stone towards a complete system for the creation and validation of complex missions to be carried out by multiple autonomous assets, and herein lies

	Offshore	MCM	Combination	Limited	All
Experimental vs. Base	0.0314	0.1994	n/a	0.0096	0.0196
Experimental vs. User	0.0282	0.2341	0.9449	0.0101	0.0200

TABLE I

THE T-TEST RESULTS FOR COMPARISONS *experimental* VS. *base* AND *experimental* VS. *user*, WITHOUT THE INCLUSION OF THE DATA FROM PARTICIPANT 13. GROUPS ARE THE SAME AS IN FIGURE 9.

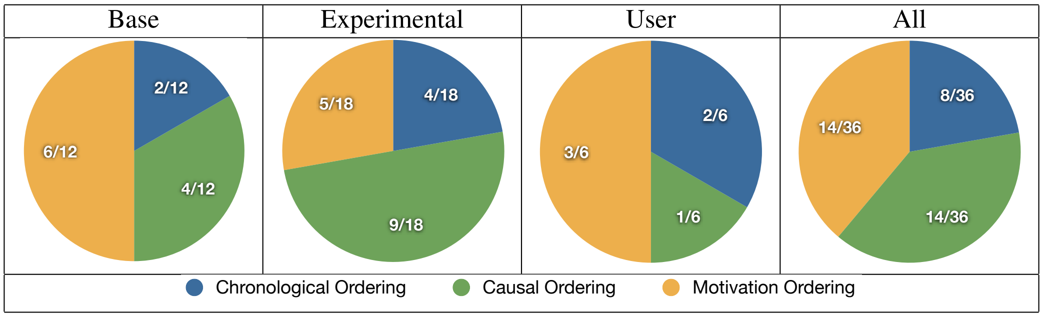

Fig. 11. Orderings used by the participants when describing the events of the offshore and MCM scenarios.

its contribution. It represents a functional prototype which shows clear potential to advance the state of the art.

The envisioned system would move user interaction forward towards a dialogue, allowing the user to participate more fully in the debrief whilst also providing the guidance and additional contextually relevant information which is lacking in the current state of the art. In other words, a system which conforms to the framework described in [15].

REFERENCES

[1] A. Cormack, D. M. Lane, and J. Wood, "Operational experiences for maritime homeland security operations," in *IEEE International Conference Oceans 2010*, May 2010.

[2] J. Evans, P. Patron, B. Privat, N. Johnson, and C. Capus, "Autotracker: Autonomous inspection – capabilities and lessons learned in offshore operations," in *Proceedings of the IEEE Oceans Conference 2009*, 2009.

[3] J. Jamieson and I. Tena, "Autonomous vehicle meets new challenges," in *Proceedings of Deep Offshore Technology*, February 2009.

[4] F. Maurelli, J. Cartwright, N. Johnson, and Y. Petillot, "Nessie IV autonomous underwater vehicle wins the SAUC-E competition," in *Proceedings of Robotica 2010*, 2010.

[5] C. Sotzing and D.M.Lane, "Improving the co-ordination efficiency of limited communication multi-AUV operations using a multi- agent architecture," *Journal of Field Robotics*, vol. 27, no. 4, pp. 412–429, July 2010.

[6] N. A. Johnson and D. Lane, "Abstracting the planner down down: An architecture for planner based control of autonomous vehicles," in *Proceedings of the 27th Workshop of the UK Planning and Scheduling Special Interest Group*, 2008.

[7] P. Patrón, D. M. Lane, and Y. Petillot, "Interoperability of agent capabilities for autonomous knowledge acquisition and decision making in unmanned platforms," in *Proceedings of the IEEE Oceans Conference 2009 - Europe*, 2009.

[8] E. Miguelaez, P. Patron, K. E. Brown, Y. R. Petillot, and D. M. Lane, "Semantic knowledge-based framework to improve the situation awareness of autonomous underwater vehicles," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 2010.

[9] SeeByte Ltd., "Seetrack military," Software. Online brochure available at http://www.seebyte.com/Products/Military/index.html. Last checked on 26/11/2009.

[10] Hydroid Inc., "Remus vehicle interface program." Software. Online brochure available at http://www.hydroid.com/software.html. Last checked 26/11/2009.

[11] Hafmynd Efh., "Gavia contol center," Software. Online brochure available at: http://www.gavia.is/products/user_software.html. Last checked on 26/11/2009.

[12] K. Collins, "Untethered AUV's can reduce costs for offshore inspection jobs," in *Proceedings of the 1993 Offshore Technology Conference*, May 1993.

[13] I. Woodrow, C. Purry, A. Mawby, and J. Goodwin, "Autonomous AUV mission planning and replanning - towards true autonomy," in *Proceedings of the 14th International Symposium on Unmanned Untethered Submersible Technolog*, August 2005.

[14] J. L. Hinton, M. Williams, and B. Kirby, "Human centred perspecivies on mission planning for autonomous systems," in *Proceedings of the 1st Annual SEAS DTC Technical Conference*. BAE Systems Advanced Technology Centre, July 2006.

[15] N. Johnson, P. Patron, and D. Lane, "The importance of trust between operator and AUV: Crossing the human/computer language barrier," in *Proceedings of the IEEE Oceans Conference 2007 - Europe*, June 2007.

[16] A. C. Graesser, K. Hauft-Smith, A. D. Cohen, and L. D. Pyles, "Advanced outlines, familiarity and text genre on retention of prose," *Journal of Experimental Education*, vol. 48, 1980.

[17] N. Pennington and R. Hastie, "Evidence evaluation in complex decision making." *Journal of Personality & Social Psychology*, vol. 51, no. 2, pp. 242–258, August 1986.

[18] ——, "Explaining the evidence: Tests of the story model for juror decision making," *Journal of Personality and Social Psychology*, vol. 62, no. 2, pp. 189–206, 1992.

[19] ——, "Reasoning in explanation-based decision making," *Cognition*, vol. 49, pp. 123–163, 1993.

[20] J. Hoffmann and B. Nebel, "The FF planning system: Fast plan generation through heuristic search," *Journal of Artificial Intelligence Research*, vol. 14, pp. 253–302, May 2001.

[21] A. C. Graesser, B. Olde, and B. Klettke, *Narrative Impact: Social and Cognitive Foundations*, 1st ed. Lawrence Erlbaum, July 2002, ch. How does the mind construct and represent stories?

[22] S. Chatman, *Story and Discourse*. Cornell University Press, 1980.

[23] R. A. Zwaan and G. A. Radvansky, "Situation models in language comprehension and memory," *Psychological Bulletin*, vol. 123, no. 2, pp. 162–185, 1998.

[24] L. V. Snyder and M. S. Daskin, "A random-key genetic algorithm for the generalized traveling salesman problem," Department of Industrial Engineering and Management Sciences, Northwestern University, Tech. Rep., 2004.

[25] E. Reiter, "SimpleNLG package," Web Page : http://www.csd.abdn.ac.uk/ ereiter/simplenlg/. Last Checked: 1/12/2008.