Dynamic Generation and Refinement of Robot Verbalization

Vittorio Perera, Sai P. Selveraj, Stephanie Rosenthal, Manuela Veloso

***Abstract*— With a growing number of robots performing autonomously without human intervention, it is difficult to understand what the robots experience along their routes during execution without looking at execution logs. Rather than looking through logs, our goal is for robots to respond to queries in natural language about what they experience and what routes they have chosen. We propose verbalization as the process of converting route experiences into natural language, and highlight the importance of varying verbalizations based on user preferences. We present our verbalization space representing different dimensions that verbalizations can be varied, and our algorithm for automatically generating them on our CoBot robot. Then we present our study of how users can request different verbalizations in dialog. Using the study data, we learn a language model to map user dialog to the verbalization space. Finally, we demonstrate the use of the learned model within a dialog system in order for any user to request information about CoBot's route experience at varying levels of detail.**

I. INTRODUCTION

We have been investigating autonomous mobile service robots for several years. Our robots perform services that involve moving between locations in our buildings, just traveling to a destination, transporting items from one place to another, or accompanying visitors to offices. Our novel and robust solutions to many challenges of such autonomous behavior have led to the autonomous navigation of more than 1,000km by the robots within the last 3-4 years [1].

Because of the success of the autonomous algorithms, our and other robots consistently move in our environments and they persistently perform tasks for us without any supervision. With robots performing more autonomous behaviors without human intervention, we do not know much about their paths and experience when they arrive at their destinations without delving into the extensive log files. In this work, we propose a new research challenge, namely how to have robots respond to queries, in natural language, about their autonomous choices including their routes taken and experienced. We are interested in ways for robots to *verbalize* (an analogy to visualization) their experiences via natural language.

We notice that different people in the environment may be interested in different specific information, for specific parts of the robot's experience, at different levels of detail, and at different times. A one-size-fits-all verbalization will not satisfy all users. For example, as robotics researchers interested in debugging our robots' behaviors, we often would like our robot to recount its entire path in great detail. On the other hand, an office worker may only want a robot to identify why it arrived late. These variances in preferences are echoed in prior literature in which autonomous systems explain their behavior [2], [3], [4].

In prior work, we have introduced *verbalization spaces* as a way to capture the fact that descriptions of the robot experience are not unique and can greatly vary in a space of different dimensions. We introduced three dimensions of our verbalization space, namely abstraction, specificity, and locality, and associate different levels to each dimension. Based on the underlying geometric map of an environment used for route planning in addition to semantic map annotations, our automated verbalization algorithm generates different explanations as a function of the desired preference within the verbalization space. We present a summary of this prior work including an example verbalization for our CoBot robot in our environment.

In this work, we pursue our research addressing the fact that people will want to request different types of verbalizations through dialog, and may even want to revise their requests through dialog as the robot verbalizes it's route experiences. We present a crowdsourced online study in which participants were told to request types of information represented in our verbalization space. We then provide the robot's verbalization response and asked the participants to write a new request to change the type of information in the presented verbalization. Using the verbalization requests collected from the study, we learn a mapping from the participant-defined language to the parameters in our verbalization space. We show the accuracy of the learned language model increases in the number of participants in our study, indicating that while the vocabulary was diverse it also converged to a manageable set of keywords with a reasonable participant sample size (100 participants). Finally, we demonstrate human-robot dialog that is enabled by our verbalization algorithm and our learned verbalization space language classifier.

II. RELATED WORK

We identify three main categories in the literature on automatically generating explanations or summaries of planned or perceived behavior: 1) intelligibility or explanation of machine learning algorithms, 2) summarizing perceived behavior, and 3) generating directions for humans to follow.

One of the main focus in Human-Computer Interaction research is developing ways for machine learning applications to *intelligibly* explain their reasoning to users (e.g., for context-aware systems [2]). The studies performed on

V. Perera and M. Veloso are with the Computer Science Department, S. Selveraj with the Robotic Institute and S. Rosenthal with Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA 15203 (e-mails: vdperera@cs.cmu.edu, spandise@andrew.cmu.edu, srpomerantz@sei.cmu.edu, mmv@cs.cmu.edu).

intelligibility focus in multiple directions. In [5], the authors look at how users can query applications for information or explanations. The focus of [6], [7] is to explore how the generated explanation can affect the users' mental model of how the applications work. Last, [8] shows how automatically generated explanation can increase the users trust. Another relevant problem is providing summaries or generating narrative of perceived behavior. This problem has been addressed in many different scenarios such as: Robocup soccer games [9], [10], wartime exercises [11], video conferencing sessions [12], or sports games [13]. Finally, automatically generating navigation instructions and dialog for people to follow and understand has become, more and more, a relevant problem in GPS applications (e.g., [14]) and robotics (e.g., [15], [3], [4]).

A common aspect of prior work is the need to vary explanations and summaries according to the user's preference. In [16] the authors show how human direction givers do not generate the same directions for every person. Recently, [3] found that navigation directions *should*: 1) provide differing levels of specificity at different locations in the route and 2) use abstract landmarks in addition to more concrete details. Although the need for parametrized summaries is well documented, none of the prior work, to our knowledge, measures those parameters and contributes an algorithm for varying them.

Our previous work on verbalization space and verbalization algorithm [17] is briefly summarized in Section III. The focus of this work is on how a user might request a variety of verbalizations. The literature of both Human-Human and Human-Robot Interaction focus on how to request additional information when the instructions provided are not clear [18], [19], [20], [21]. Our approach differs as, rather than focusing on changing or repairing instructions when there is a communication breakdown, we allow users to proactively request language variation based on preferences. The contribution of our experiment is twofolds, first to understand the user's language for specifying what information they would want in a verbalization, and second to understand user's language to *change* a verbalization to receive new or different information. We then create a predictive models and demonstrate how we can use them to predict the verbalization preference.

III. ROUTE VERBALIZATION

Previously, we have defined *verbalization* as the process by which an autonomous robot converts its own experience into language. We represent the variations in possible explanations for the same robot experience in the *verbalization space* (VS). Each region in verbalization space represents a different way to generate explanations to describe a robot's experience by providing different information as preferred by the user. Specifically, given an annotated map of the environment, a route plan through the environment, and a point in our verbalization space, our Variable Verbalization Algorithm generates a set of sentences describing the robot's experience following the route plan. We summarize each of these aspects in turn and then provide example verbalizations for our indoor mobile robot CoBot.

A. Environment Map and Route Plans

Our robot maintains an environment map with semantic annotations representing high level landmarks of interest. We define the map $M = \langle P, E \rangle$ as set of points $p = (x, y, m) \in P$ representing unique locations (x, y) locations for each floor map m, and edges $e = \langle p_1, p_2, d, t \rangle \in E$ that connect two points p_1, p_2 taking time t to traverse distance d.

The map is annotated with semantic *landmarks* represented as room numbers (e.g., 7412, 3201) and room type (office, kitchen, bathroom, elevator, stairs, other). The map is also annotated with lists of points as *corridors* which typically contain offices (e.g., "7400 corridor" contains (office 7401, office 7402, ...)) and *bridges* as hallways between offices (e.g., "7th floor bridge" contains (other 71, other 72)).

Using our map, a route planner produces route plans as trajectories through our map. The route plan is composed of a starting point S, finish point F, an ordered list of intermediate waypoints $W \subset P$, and a subset of edges in E that connect S to F through W. Our route planner annotates route plans with *turning points* (e.g., [22]) to indicate the locations where the robot turns after moving straight.

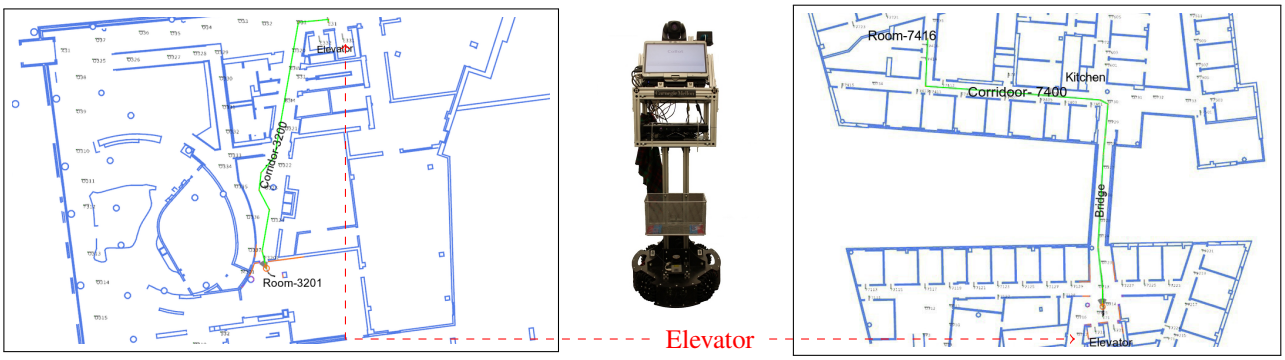
B. Verbalization Space Components

For any given route plan, many different verbalization summaries can be generated. We formalize the space of possible verbalizations as the *verbalization space* (VS) consisting of a set of axes or parameters along which the variability in the explanations are created. For the purpose of describing the path of the CoBot, our VS contains three orthogonal parameters with respect to the environment map and route plan – abstraction, locality, and specificity. These parameters are well-documented in research, though they are not exhaustive ([2], [3], [4]).

Abstraction A: Our abstraction parameter represents the vocabulary or corpus used in the text generation. In the most concrete form (Level 1), we generate explanations in terms of the robot's world representation, directly using points (x, y, m) in the path. Our Level 2 derives angles, traversal time and distances from the points used in Level 1. Level 3 abstracts the angles and distances into right/left turns and straight segments. And finally at the highest level of abstraction, Level 4 contains location information in terms of landmarks, corridors, and bridges from our annotated map.

Locality L: Locality describes the segment(s) of the route plan that the user is interested in. In the most general case, the user is interested in the plan through the entire Global Environment. They may only be interested in a particular Region defined as a subset of points in our map (e.g., the 8th floor or Building 2), or only interested in the details around a Location (e.g., 8th floor kitchen or office 4002).

Specificity S: Specificity indicates the number of concepts or details to discuss in the text. We reason about three levels of specificity, the General Picture, the Summary, and the Detailed Narrative. The General Picture contains a short



3rd Floor map 7th Floor map

Fig. 1. Example of our mobile robot's planning through our buildings. Building walls are blue, the path is green, the elevator that connects the floors is shown in red and shown in black text are our annotations of the important landmarks.

description, only specifying the start and end points or landmarks, the total distance covered and the time taken. The Summary contains more information regarding the path than General Picture, and the Detailed Narrative contains a complete description of the route plan in the desired locality, including a sentence between every pair of turning points.

C. Variable Verbalization Algorithm

Given the route plan, the verbalization preference in terms of (A, L, S), and the environment map, our Variable Verbalization (VV) Algorithm translates the robot's route plan into plain English (pseudocode in Algorithm 1). We demonstrate algorithm with an example CoBot route plan from starting point "office 3201" to finish point "office 7416" as shown in Figure 1. In this example, the user preference is (Level 4, Global Environment, Detailed Narrative).

Algorithm 1 Variable Verbalization (VV) Algorithm

Input: *path, verb_pref, map* **Output:** *narrative*

//The verbalization space preferences
1: $(a, l, s) \leftarrow$ *verb_pref*
//Choose which abstraction vocabulary to use
2: *corpus* $\leftarrow$ ChooseAbstractionCorpus(a)
//Annotate the path with relevant map landmarks
3: *annotated_path* $\leftarrow$ AnnotatePath(*path, map*, a)
//Subset the path based on preferred locality
4: *subset_path* $\leftarrow$ SubsetPath(*annotated_path,* l)
//Divide the path into segments, one per utterance
5: *path_segments* $\leftarrow$ SegmentPath(*subset_path, s*)
//Generate utterances for each segment
6: *utterances* $\leftarrow$ NarratePath(*path_segments, corpus, a, s*)
//Combine utterances into full narrative
7: *narrative* $\leftarrow$ FormSentences(*utterances*)

The VV Algorithm first uses abstraction preference a to choose which corpus (points, distances, or landmarks) to use when generating utterances (Line 2). Since the abstraction preference in the example is Level 4, the VV algorithm chooses corpus of landmarks, bridges and corridors from the annotated map. The VV algorithm then annotates the route plan by labeling the points along the straight trajectories by their corridor or bridge name and the route plan turning points based on the nearest room name.

Once the path is annotated with relevant locations, the algorithm then extracts the subset of the path that is designated as relevant by the locality preference l (Line 4). In this case, the locality is Global Environment and the algorithm uses the entire path as the subset. The VV algorithm then determines the important segments in the path to narrate with respect to the specificity preference s (Line 5). For Detailed Narratives, our algorithm uses edges between all turning points, resulting in descriptions of the corridors and bridges, landmarks, and the start and finish points:

{s1: Office 3201, s2: Corridor 3200, s3: Elevator,
s4: 7th Floor Bridge, s5: 7th Floor Kitchen,
s6: Corridor 7400, s7: Office 7416}

The VV Algorithm then uses segment descriptions and phrase templates to compose the verbalization into English utterances (Line 6). Each utterance template consists of a noun N, verb V, and route plan segment description D to allow the robot to consistently describe the starting and finish points, corridors, bridges, landmarks, as well as the time it took to traverse the path segments. The templates could also be varied, for example, to prevent repetition by replacing the verbs with a synonym (e.g., [10]). The following are the templates used on CoBot for the Level 4 abstractions. We note next to the D whether the type of landmark is specific (e.g., the template must be filled in by a corridor, bridge, etc), and we note with a slash that the choice of verb is random.

- "[I]$_N$ [visited/passed]$_V$ the [___]$_{D:room}$"
- "[I]$_N$ [took]$_V$ the elevator and went to the [___]$_{D:floor}$"
- "[I]$_N$ [went through/took]$_V$ the [___]$_{D:corridor/bridge}$"
- "[I]$_N$ [started from]$_V$ the [___]$_{D:start}$"
- "[I]$_N$ [reached]$_V$ [___]$_{D:finish}$"

The template utterances are joined using "then"s but could also be kept as separate sentences. Using the filled-in templates, the VV Algorithm generates the following verbalization (Line 7):

I started from office 3201, I went through the 3200 corridor, then I took the elevator and went to the seventh floor, then I took the 7th floor bridge, then I passed the 7th floor kitchen, then I went through the 7400 corridor, then I reached office 7416.

IV. DIALOG TO REVISE VERBALIZATIONS

Our Variable Verbalization Algorithm takes as input a user's explanation request (a, l, s) in terms of level a of Abstraction, l of Locality, and s of Specificity. We further envision the user to engage in a dialog with the robot to incrementally revise their verbalization preferences. In this section, we contribute an approach for mapping the user's dialog onto a verbalization preference, along the dimensions of the Verbalization Space (VS).

As an example, consider the following request to the robot for an explanation: "Please, tell me exactly your experience for your whole path to get here." Since this sentence refers to the "whole path," the robot uses the Global Environment level in the Locality dimension of the Verbalization Space. Furthermore, as the user uses the term "exactly," the explanation should be at the level of Detailed Narrative in the dimension of Specificity. Finally, although no language feature in the request directly refers to a level of Abstraction, the robot may use a high level of Abstraction, as its default. We then concretely address the problem of dialoguing with the robot to revise an explanation. Once a user asks for and receives a route verbalization, they could be interested in refining such description. If we continue the above example, after the robot offers a detailed description of its path, the user could ask: "OK robot, now tell me only what happened near the elevator." The user is hence asking a revised summary of the task executed, where the language should map the explanation to the same values for Abstraction and Specificity as in the initial description, but now focusing the Locality on the region of the elevator.

Our learned mapping from language-based requests to points in the verbalization space allows the user to dynamically refine previous preferences through dialog.

A. Data Collection

In order to enable a robot to correctly infer the user's initial VS preferences as well as how to move in the VS to refine the preferences, we gathered a corpus of 2400 commands (available at *link*) from a total of 100 participants through an Amazon Mechanical Turk survey (www.mturk.com) in which each participant was asked 12 times to request information about our robot's paths and then refine their request for different information. Table I shows a sample of the corpus.

Please give me a summary of statistics regarding the time that you took in each segment.
Can you tell me about your path just before, during, and after you went on the elevator?
How did you get here?
Can you please eliminate the time and office numbers?
What is the easiest way you have to explain how you came to my office today?
Robot, can you please further elaborate on your path and give me a little more detail?

TABLE I

SAMPLE SENTENCES FROM THE CORPUS

After giving consent to partake in the survey, the users were given instructions in order to complete the survey. These instructions included: 1) a short description of the robot capabilities (i.e., execute task for users and navigate autonomously in the environment) and 2) the context of the interaction with the robot. In particular, we asked the users to imagine the robot had just arrived at their office and they were interested in knowing how it got there. Each time the robot arrived at their office, the participants were given:

- A free-response text field to enter a sentence requesting a particular type of summary of the robot's path,
- An example of the summary the robot could provide, and finally
- A second free-response text field to enter a new way to query the robot assuming their interest changed.

This process was repeated 12 times for different parts of our VS. Figure 2 shows the first page of the survey.

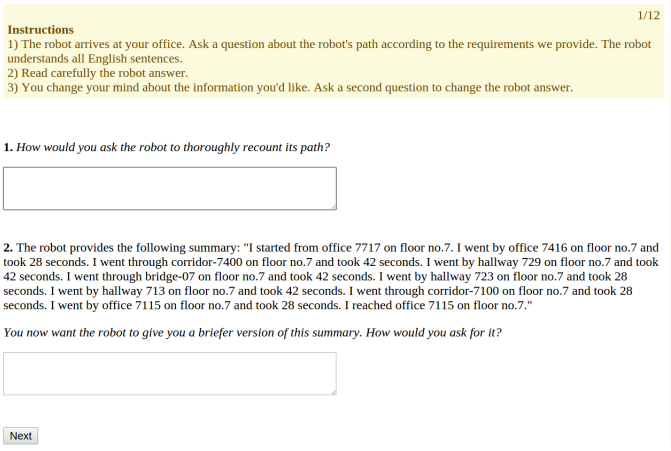


Fig. 2. The survey used to gather out the data corpus. The instructions above the two text fields read: "How would you ask the robot to thoroughly recount its path" and "You now want the robot to give a briefer version of this summary. How would you ask for it?"

We note that the instructions to our survey purposefully did not mention the concept of verbalization and did not introduce any of the three dimensions of the verbalization space. Users hence were not primed to use specific ways to query the robot. However, as the sentences in our corpus should cover the whole verbalization space, when asking for the initial sentence on each page, we phrased our request in a way that would refer to a point on one of the axis of the VS. As an example, in Figure 2, we ask for a sentence matching a point with Detailed Narrative Specificity, and therefore we ask "How would you ask the robot to thoroughly recount its path?". The second sentence we requested on each page refers to a point on the same axis but with opposite value. In Figure 2, we look for a sentence matching a point with General Picture specificity, and we ask the user "You now want the robot to give you a briefer version of this summary. How would you ask for it?". In the first 6 pages of the survey, we asked for an initial sentence matching a point for each possible dimension (Abstraction/Specificity/Locality) at the extreme values. The same questions were asked a second time in the remaining 6 pages of the survey. Table II shows the phrasing for each dimension/value pair.

Abstraction	High	"How would you ask the robot for an easy to read recount of its path?"
	Low	"How would you ask the robot for a recount of its path in terms of what the robot computes?"
Specificity	High	"How would you ask the robot to thoroughly recount its path?"
	Low	" How would you ask the robot to briefly recount its path?"
Locality	High	"How would you ask the robot to focus its recounting of the path near the elevator?"
	Low	"How would you ask the robot to recount each part of its entire path?"

TABLE II
PHRASING OF SURVEY INSTRUCTIONS

V. LEARNING DIALOG MAPPINGS

We frame the problem of mapping user dialog to VS dimensions of Abstraction, Specificity and Locality as a problem of text classification. In particular, we consider the six possible labels corresponding to two levels, high or low extremes, for each of the three axes of the verbalization space. The corpus gathered from the Mechanical Turk survey was minimally edited to remove minor typos (e.g., 'pleaes' instead of 'please') and automatically labeled. The automatic labeling of the corpus was possible since the ground truth was derived directly from the structure of the survey itself.

To perform the classification, we tried several combinations of features and algorithms. Here, we report on the most successful ones. The features considered for our classification task are unigrams, both in their surface and lemmatized form, bigram and word frequency vectors. We also considered two different algorithms, a Naive Bayes Classifier and Linear Regression. Figure 3 shows the results.

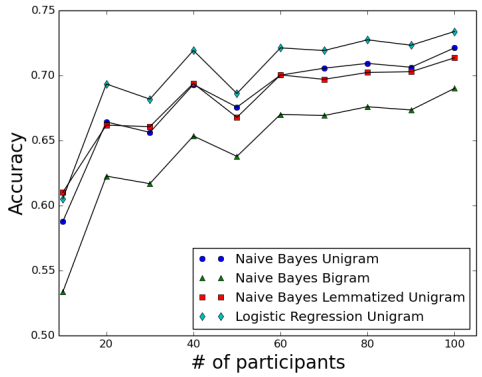


Fig. 3. Experimental results. On the X axis the number of users used to train and test the model, on the Y axis the accuracy achieved.

The X axis shows the number of participants, randomly selected from the pool of 100 survey takers, used to train the model. The Y axis, shows the average accuracy over 10 leave-one-out cross validation tests. As the number of participants increases, all of the proposed approaches improve in performance, as the size of the corpus increases proportionally. Once a robot is deployed and is able to gather more and more sentences asking to verbalize a path, it will then further improve the accuracy of the classification.

When trained on the whole corpus, Logistic Regression achieves the best results with 73.37% accuracy. The accuracy for the Naive Bayes Classifer is 72.11%, 71.35%, and 69% when trained using unigrams, lemmatized unigrams, and bigrams, respectively. Interestingly, the Bayes Classifier and Linear Regression perform similarly with data points differing by less than 2%. Lemmatizing the unigrams does not appear to have a strong effect on the classifier accuracy. Finally, using bigrams negatively effects the classification accuracy. While bigrams encode more information than unigrams, bigrams also naturally produce a more sparse representation of the sentence, which coupled with the size of our corpus, leads to lower accuracy rates.

All the models were trained on the whole corpus and therefore their use is twofold. First, we can use them to infer the correct point in the VS when a user initially queries the robot. Second, we can use the models to dynamically update these values during the dialog. Since our best model achieves 73.37% accuracy the robot is not always going to provide the correct level of information when verbalizing its task. This should not be considered a problem as, using the same models enables the iterative query refinement through dialog.

A. Demonstration

We demonstrate how the learned dialog mapping guides the verbalization process. In this example, we used the Naive Bayes Classifier trained using only unigrams as features to predict the labels of the sentences. We start with the initial sentence given by the user:

USER: Please tell me exactly how you got here.

Our classifier finds the most likely label for the sentence Detailed Narrative, corresponding to point 1 in Figure 4. Accordingly, the robot describes its task:

ROBOT: I started at office 7717. I went by office 7416 and took 28 seconds. I went through corridor 7400 and took 42 seconds. I went by the open area 730 and took 28 seconds. I reached office 7501.

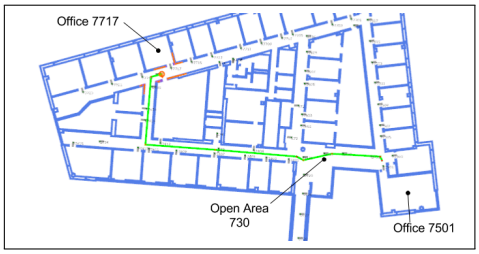
USER: OK, now only tell me about the open area.

Once again the robot classifies the input sentence returning the label Locality Location-specific. Our algorithm now updates the parameters of the verbalization algorithm by increasing the value of locality to produce a new summary, now corresponding to the transitionfrom point 1 to 2 in Figure 4.

ROBOT: Near open area 730. I started from office 7407 on floor and went through corridor-7400.

USER: Can you rephrase in terms of what you computed?

This final command is classified as Level 1 Abstraction. The parameters of the algorithm are updated as shown in the transition from 2 to 3 in Figure 4 and the robot produces a third summary of its task.



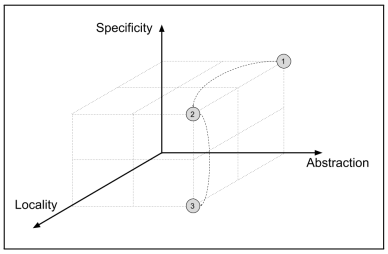


Fig. 4. Demonstration of movements in the verbalization space. On the left the path the robot is going to verbalize. On the right three different point in the verbalization space.

ROBOT: Near O730. Start point is (-1.36, 31.22) on floor GHC7, end point is (10.79, 30.20) on floor GHC7. Start point is (10.79, 30.20) on floor no.GHC7, end point is (16.0, 30.63) on floor no.GHC7.

This dialog sequence demonstrates how the verbalization algorithm, coupled with text classification, allows the robot to understand natural language commands to verbalize its task at different levels. Moreover the user is able to specify what type of information the robot is going to provide through an interactive dialog, to further refine this information without delving into execution logs.

B. User Study

Finally, to further demonstrate our approach, we run a user study involving 7 participants. Each participant was first introduced to our CoBot robot and its capabilities, namely to autonomously navigate in the environment and perform item transportation or people guiding tasks. Next, we explained the concept of verbalization space and its three dimensions, Abstraction, Locality, and Specificity. Last, we pointed out that goal of the study was to evaluate the robot's ability to properly explain the path it had traversed. After this explanation phase, the participants were given an initial verbalization of the robot's path. This verbalization was generated by randomly selecting a point in the VS. The subjects were then instructed to provide a sentence to revise the explanation, such that the verbalization would move in a specific direction along one of the three dimensions. The robot then provided a new verbalization by applying the learned classifier, and the users were asked if the revision provided matched, did not match, or almost matched their expectations. Each user dialoged about 4 paths of the robot, and the dialog was repeated 3 times, each for each direction of the VS. There was hence a total of 84 different exchanges between a user and the robot, which were logged.

We first analyze the accuracy of the classifier, in terms of the desired dimension and direction corresponding to the language input. Even if the training from the collected corpus is clearly still limited, the classifier was correct in 54.76% of the cases, i.e., in 46 out of the 84 completely new requests given by the users. In 82,6% of these interactions, the users found that the new verbalization provided matched or almost matched their expectations. In the second step of our analysis, we looked at the remaining 38 interactions where the label returned by the classifier did not match the instructions provided. Surprisingly, the users reported that the verbalization matched their expectations in 21.05% of the cases. By a closer inspection, we found out that the users were confused with the instructions and did not match the directions and dimensions in the verbalization space. For instance, when asked to provide a sentence to move the verbalization towards a *lower specificity* (i.e., a shorter description), one of the users asked "Tell me about your whole path." The classifier labeled this sentence as low locality, the robot extended the verbalization, previously limited in the surroundings of an elevator, to the entirety of the path and matched the users expectations. Table III summarizes the results of the users study.

	Match	Almost Match	Don't Match	
Correct Label	32	6	8	46
Incorrect Label	8	6	24	38
				84

TABLE III
RESULTS OF THE USERS STUDY.

In conclusion, if we consider both the cases where the classifier returned the label meant in the instructions and the cases where the users considered the new verbalization to match their expectations, the dialog was able to provide a correct verbalization in 64.28% of the cases.

VI. CONCLUSIONS

A significant challenge with autonomous mobile robots is understanding what they are doing when there is no human around. We propose verbalization as the process of converting sensor data into natural language to describe a robot's experiences. We review our verbalization space representing different dimensions that verbalizations can be varied, and our algorithm for automatically generating them on our CoBot robot. Then we present our study of how users can request different verbalizations in dialog. Using 2400 utterances collected from the study, we demonstrate that it is possible to learn a language model that maps user dialog to our verbalization space. With greater than 70% accuracy, a robot that uses this model can predict what verbalization a person expects and refine the prediction further through continued dialog. We demonstrate this ability with example verbalizations for CoBot's route experiences.

ACKNOWLEDGMENTS

This material is based upon work partially funded by a grant from the Future of Life Institute, by NSF under grant number IIS-1012733; and ONR under grant number N00014-09-1-1031. In addition it was funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN AS-IS BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT. [Distribution Statement A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution. DM-0003431

REFERENCES

[1] M. Veloso, J. Biswas, B. Coltin, and S. Rosenthal, "CoBots: Robust Symbiotic Autonomous Mobile Service Robots," in *Proceedings of IJCAI'15*, July 2015.

[2] A. K. Dey, "Explanations in context-aware systems." in *ExaCt*, 2009, pp. 84–93.

[3] D. Bohus, C. W. Saw, and E. Horvitz, "Directions robot: In-the-wild experiences and lessons learned," in *Proceedings of the 2014 international conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 637–644.

[4] J. Thomason, S. Zhang, R. Mooney, and P. Stone, "Learning to interpret natural language commands through human-robot dialog," in *Proceedings of IJCAI'15*, July 2015.

[5] B. Y. Lim, A. K. Dey, and D. Avrahami, "Why and why not explanations improve the intelligibility of context-aware intelligent systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 2119–2128.

[6] T. Kulesza, S. Stumpf, M. Burnett, and I. Kwan, "Tell me more?: the effects of mental model soundness on personalizing an intelligent agent," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1–10.

[7] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W.-K. Wong, "Too much, too little, or just right? ways explanations impact end users' mental models," in *Visual Languages and Human-Centric Computing (VL/HCC), 2013 IEEE Symposium on*. IEEE, 2013, pp. 3–10.

[8] A. Bussone, S. Stumpf, and D. O'Sullivan, "The role of explanations on trust and reliance in clinical decision support systems," *IEEE International Conference on Healthcare Informatics*, 2015.

[9] D. Voelz, E. André, G. Herzog, and T. Rist, "Rocco: A robocup soccer commentator system," in *RoboCup-98: Robot Soccer World Cup II*. Springer, 1999, pp. 50–60.

[10] M. Veloso, N. Armstrong-Crews, S. Chernova, E. Crawford, C. McMillen, M. Roth, D. Vail, and S. Zickler, "A team of humanoid game commentators," *International Journal of Humanoid Robotics*, vol. 5, no. 03, pp. 457–480, 2008.

[11] L. J. Luotsinen, H. Fernlund, and L. Bölöni, "Automatic annotation of team actions in observations of embodied agents," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007, p. 9.

[12] A. Yengui and N. Mahmoud, "Semantic annotation formalism of video-conferencing documents," *Cognition and Exploratory Learning in Digital Age (CELDA 2009), Rome (Italy)*, 2009.

[13] N. D. Allen, J. R. Templon, P. S. McNally, L. Birnbaum, and K. J. Hammond, "Statsmonkey: A data-driven sports narrative writer." in *AAAI Fall Symposium: Computational Models of Narrative*, 2010.

[14] R. Belvin, R. Burns, and C. Hein, "Development of the hrl route navigation dialogue system," in *Proceedings of the first international conference on Human language technology research*. Association for Computational Linguistics, 2001, pp. 1–5.

[15] R. Kirby, F. Broz, J. Forlizzi, M. P. Michalowski, A. Mundell, S. Rosenthal, B. P. Sellner, R. Simmons , K. Snipes, A. Schultz, and J. Wang, "Designing robots for long-term social interaction," in *Proceedings of IROS'05)*. IEEE, August 2005, pp. 2199 – 2204.

[16] A. MacFadden, L. Elias, and D. Saucier, "Males and females scan maps similarly, but give directions differently," *Brain and Cognition*, vol. 53, no. 2, pp. 297–300, 2003.

[17] S. Rosenthal, S. P. Selvaraj, and M. Veloso, "Verbalization: Narration of Autonomous Mobile Robot Experience," in *Proceedings of IJCAI'16, the 26th International Joint Conference on Artificial Intelligence*, New York City, NY, July 2016.

[18] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz, "Enabling effective human-robot interaction using perspective-taking in robots," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 35, no. 4, pp. 460–470, 2005.

[19] C. Torrey, A. Powers, M. Marge, S. R. Fussell, and S. Kiesler, "Effects of adaptive robot dialogue on information exchange and social relations," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 126–133.

[20] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski, "Gracefully mitigating breakdowns in robotic services," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 203–210.

[21] B. M. Allison Saupp, "Effective task training strategies for human and robot instructors," *Autonomous Robots*, vol. 39, pp. 313–329, 2015.

[22] J. Biswas and M. Veloso, "WiFi Localization and Navigation for Autonomous Indoor Mobile Robots," in *Proceedings of ICRA'10*, Anchorage, AL, May 2010.