On Natural Language Dialogue with Assistive Robots

Vladimir A. Kulyukin
Computer Science Assistive Technology Laboratory
Department of Computer Science
Utah State University
vladimir.kulyukin@usu.edu

ABSTRACT
This paper examines the appropriateness of natural language dialogue (NLD) with assistive robots. Assistive robots are defined in terms of an existing human-robot interaction taxonomy. A decision support procedure is outlined for assistive technology researchers and practitioners to evaluate the appropriateness of NLD in assistive robots. Several conjectures are made on when NLD may be appropriate as a human-robot interaction mode.

Categories and Subject Descriptors

General Terms
Performance, Design, Experimentation, Human Factors.

Keywords
assistive technology, natural language dialogue, assistive robotics.

1. INTRODUCTION
Human-robot interaction (HRI) is an active research area whose findings are of significance to assistive robotics. Robotic solutions have now been implemented in wheelchair navigation [45], hospital delivery [40], microsurgery [44], robot-assisted navigation [1, 6, 27, 28], care for the elderly [34], and life support partners [24]. In each of these tasks, the effectiveness of HRI is critical.

An important question facing many assistive technology (AT) researchers and practitioners today is what is the best way for humans to interact with assistive robots. As more robotic devices find their way into healthcare and rehabilitation, it is imperative to find ways for humans to effectively interact with those devices. Many researchers have argued that natural language dialogue (NLD) is a promising HRI mode [12, 25, 32, 40, 42]. In this paper, this claim is examined in the context of assistive robotics. In examining the NLD claim, we do not argue either for or against NLD per se. Rather, our objective is to develop a decision support procedure that may be helpful to the assistive robotics community and the AT community at large in evaluating the appropriateness of NLD in prospective assistive robotic applications.

Our paper is organized as follows. First, assistive robots are defined in terms of an existing HRI taxonomy. Second, the NLD claim is stated and the main arguments for and against it are examined. Third, the NLD claim is critiqued in the context of assistive robotics. Fourth, a decision support procedure is outlined for evaluating the appropriateness of NLD in assistive robots. Finally, several conjectures are made on when NLD may be appropriate as an HRI mode.

2. WHAT IS AN ASSISTIVE ROBOT?
Before we can discuss the appropriateness of NLD in assistive robots, we need to define what is meant by an assistive robot. One way to define a concept is to place that concept in an existing taxonomy. We will use the HRI taxonomy developed by Yanco and Drury [46].

Yanco and Drury [46] postulate eight categories: 1) autonomy level, 2) amount of intervention, 3) human robot ratio, 4) type of shared human robot interaction, 5) decision support for operators, 6) criticality, 7) time/space, and 8) composition of robot teams. The autonomy level measures the percentage of time that the robot operates independently. The amount of intervention measures the amount of time that the robot’s operator operates the robot. Both categories are real-valued in the range from 0 to 1.

The human robot ratio is a non-reduced fraction of the number of human operators over the number of robots. The type of shared human robot interaction elaborates the human robot ratio through a finite set of symbolic values, e.g., one human - one robot, one human - robot team, one human - multiple robots, human team - one robot, etc. The decision support category defines information provided for operators for decision support. This category has the following symbolically valued subcategories: available sensors, available sensor information, type of sensor fusion, and type of pre-processing.

Criticality estimates how critical it is for a robot to complete a specific task. A critical task is defined to be one where a failure affects the life of a human. Criticality can be high, medium, or low. The time/space category describes HRI based on whether humans and robots interact in the same time (synchronous) or not (asynchronous) and at the same space.
Finally, the composition of robot teams can be homogenous or heterogenous. If the robot team is heterogeneous, the available robot types can be further specified.

Where in this taxonomy can we place assistive robots? To answer this question, we first need to determine which categories should be used. We believe that two categories in the proposed taxonomy may not be appropriate for our purposes: decision support and composition of robot teams. Our main concern with the decision support category is that it is underspecified: the types of available sensors and sensor fusion algorithms used in assistive robots vary greatly, and these can be classified only at a very specific level. Moreover, since sensor data pre-processing is determined by sensor fusion algorithms, it cannot be given an independent value. Although composition of robot teams is well defined, it may not be applicable to assistive robots for practical reasons: we are not aware of a single AT system in which a team of assistive robots is used. This is not to say, of course, that such teams are impossible in principle and will not be deployed in the future.

Let us now attempt to describe assistive robots in terms of the remaining six categories. Since an assistive robot is designed to provide support to a disabled individual [6, 28, 45] or to a caregiver [40, 44], it is unlikely to be fully autonomous. More often than not, it is designed to receive directions from a human, complete the tasks specified by those directions, report on its progress, and wait for more directions. Thus, its autonomy level is less than 1 and the amount of human intervention is greater than 0. The majority of existing assistive robots have a human-robot ratio of 1: one human operating one robot. For some assistive robots, e.g., Hygeirobot, a hospital delivery robot [40], the ratio can be more than 1. In principle, many nurses can request Hygeirobot to deliver certain medications to different rooms. However, this type of ratio is an exception, not the rule. Thus, the types of HRI in assistive robots are either one human - one robot (more frequent) or human team - one robot (rare). The criticality of assistive robots is high for obvious reasons. Finally, since assistive robots share the time and space with their human operators, HRI with assistive robots is synchronous and collocated.

3. The NLD Claim

The essence of the NLD claim, stated simply, is as follows: NLD is a promising HRI mode in robots capable of collaborating with humans in dynamic and complex environments [32, 40, 42]. The claim is often challenged on methodological grounds: if there is a library of robust routines that a robot can execute, why bother with NLD at all? Why not create a graphical user interface (GUI) to that library that allows human operators to invoke various routines by pointing and clicking? Indeed, many current approaches to HRI are GUI-based [10, 15, 36]. The robot's abilities are expressed in a GUI through graphical components. To cause the robot to take an action, an operator sends a message to the robot through a GUI event. The robot sends feedback to the GUI and the interaction cycle repeats.

The NLD proponents respond to this challenge with three arguments. First, it is argued that, to humans, language is the most readily available means of communication. If the robot can do NLD, the human can interact with the robot naturally, without any of the possibly steep learning curves required for many GUI-based approaches. Second, it is argued that, from the practical point of view, GUI-based approaches are appropriate in environments where the operator has access to a monitor, a keyboard, or some other hardware device, e.g., an engineer monitoring a mining robot or an astronaut monitoring a space shuttle arm. In some environments, however, access to hardware devices is either not available or impractical. For example, an injured person in an urban disaster area is unlikely to communicate with a rescue robot through a GUI. Thus, NLD in robots interacting with humans has practical benefits. Finally, an argument is made that building robots capable of NLD yields insights into human cognition [19, 25, 42]. Unlike the first two arguments made by the NLD camp, this argument does not address either the practicality or the naturalness of NLD, the most important objective being a cognitively plausible integration of language, perception, and action. Although one may question the appropriateness or feasibility of the third argument in a specific context, the argument, in and of itself, appears to be sound insomuch as the objective is to test a computational realization of some cognitive theory. Consequently, below we will focus only on the first two arguments.

Regardless of which of the three arguments they advance in defense of NLD, many researchers appear to show a consensus on what the generic architecture of the NLD system should be. The consensus architecture is shown in Figure 1. While the names of individual components vary from system to system, their overall functionality is the same or similar in all systems. The differences that exist among NLD systems have to do with how different components are realized, communication protocols among the components, and the overall distribution of functionality among the components.

The Speech Recognition module converts phonemes to symbols. Symbols are given to the natural language processing (NLP) module, which performs their syntactic and/or semantic analysis and gives the computed representation to the dialogue management system (DMS).

The DMS makes decisions as to what action should be performed by the robot. If the robot is to perform an action, a representation of that action is given to the robot hardware for execution. If the input from the NLP component requires a response, the DMS
Certain physical disabilities, e.g. spinal cord injuries, while not have any incentives to undergo the necessary training? Certain sufficient cognitive and physical abilities and, if they do, do they challenge to the AT community: do the target users have robot. The acquisition of the shared vocabulary poses another differently, the human must know the vocabulary used by the refer to that context through a shared vocabulary [16]. To put it Since NLD always happens in a context, the robot and the user sounds as directives to go to different locations. Several times the ASR erroneously recognized as route directives, which caused the some phrases said by the guided person to the interlocutor were with someone [29]. Since speech recognition runs continuously, discovered that many speech recognition errors occur when the robotic guide, we thought of using ASR as a means for the robot hardware, waits for the response, translates the response into an appropriate representation and gives that representation to the NLG module. Finally, the NLG component converts the input representation into symbols that are given to the Speech Synthesis module for vocalization.

4. A CRITIQUE OF THE NLD CLAIM
Before we offer our critique of the NLD claim, we would like to reiterate that the critique should be taken only in the context of assistive robotics. In examining the NLD claim and its arguments, we will start with the computational realization of the claim in the consensus architecture and, after identifying its weaknesses, we will analyze what impact those weaknesses have on the arguments in defense of the NLD claim.

We believe that there are three main weaknesses in the NLD architecture: the uni-directional link between speech recognition and NLP, the uni-directional link between NLG and speech synthesis, and the bi-directional link between the DMS and the robot hardware. In Figure 1, the weak links are labeled with question mark icons.

4.1 Speech Recognition
From the perspective of assistive robotics, there are two problems posed by speech recognition: speech recognition errors and shared vocabulary. Speech recognition errors present a serious safety concern for AT applications. An automatic speech recognition (ASR) system, such as Microsoft's SAPI 5.1 [3] or IBM's ViaVoice [2], may well average 95 to 97 percent accuracy in dictation experiments, where user training is available and the consequences of misrecognized words are easily absorbed. However, an assistive robot that misrecognizes 5 out of 100 commands is a definite risk, because high criticality is not guaranteed.

We learned this criticality lesson from our experiments with a robotic guide for the blind [28]. When we first started working on the robotic guide, we thought of using ASR as a means for the visually impaired to interact with the robot [41]. We quickly discovered that many speech recognition errors occur when the person guided by the robot stops and engages in conversation with someone [29]. Since speech recognition runs continuously, some phrases said by the guided person to the interlocutor were erroneously recognized as route directives, which caused the robot to start moving in a wrong direction. Several times the ASR system on our robotic guide interpreted coughs and throat clearing sounds as directives to go to different locations.

Since NLD always happens in a context, the robot and the user refer to that context through a shared vocabulary [16]. To put it differently, the human must know the vocabulary used by the robot. The acquisition of the shared vocabulary poses another challenge to the AT community: do the target users have sufficient cognitive and physical abilities and, if they do, do they have any incentives to undergo the necessary training? Certain cognitive disabilities rule out the use of speech from the start. Certain physical disabilities, e.g. spinal cord injuries, while not affecting the cognitive adequacy of the disabled, may lead to pronounced speech defects that ASR cannot handle. Sometimes training may not be feasible, because it is difficult, if not impossible, to find a representative sample of the target users. For example, it is difficult to collect a representative sample from the population of the visually impaired users of a robotic guide deployed at an airport [26]. In other contexts, the incentives for the target user to undergo training may be a serious issue. For example, a computer savvy nurse may prefer a GUI-based interface to a hospital delivery robot simply because she is already more comfortable with GUIs and must use them anyway. The same observation is applicable to many office delivery robots [32]. In a typical office environment, many office workers have access to personal computers. If there is one thing that these people know how to do, it is pointing and clicking. Thus, it may well be more convenient for them to request the robot to bring them cups of coffee through a simple GUI client on their desktops than through an NLD with the robot.

4.2 Speech Synthesis
Speech remains the primary output mode in NLD systems [38]. However, as recent research in environmental sonification shows, speech beacons have drawbacks. Tran, Letowski, and Abouchacra [43] show experimentally that speech beacons are harder to localize than non-speech beacons. Kulykuk et al. [29] suggest that visually impaired users do not always choose speech beacons when presented with an opportunity to select from a set of speech and non-speech beacons to specify a navigation event. For example, in one of our experiments, several visually impaired users opted for a non-speech beacon that consisted of a short audio file with the sound of water bubbles to signify the presence of a water fountain.

Since non-speech beacons remain a relatively new research area, unanswered questions abound. One definite advantage of speech beacons is that they do not need to be learned, whereas non-speech beacons must be learned by the user prior to using the system. On the other hand, non-speech beacons appear to require less cognitive processing than speech beacons [43] and are easier to perceive in noisy environments [29, 41]. It is also harder to engage in a dialogue with a third party when one has to attend to frequent speech beacons produced by the system [39].

Finally, there appears to be a dearth of statistically valid data on the perception of speech vs. non-speech beacons. While small samples of participants in published audio perception studies are easily explained through funding constraints, the absence of a sound experimental design framework for evaluating audio perception that is accepted by most HRI researchers is a conceptual and practical hurdle.

4.3 DMS and Robot Hardware
It is often stated that NLD-capable robots are intended to be used by people with little or no computing experience [40]. Even if one is to assume that the problems with shared vocabulary, speech recognition, and speech perception are surmountable in a particular application, hardware reliability remains a serious concern. For the NLD architecture to operate at human rates, the DMS system must be aware of the state of the robot hardware at all times. The degree of self awareness is conditional on the degree of hardware reliability. The less reliable the robot
Lack of self awareness stems from the NLD design practices prevalent in HRI. The first design practice, adopted in Hygeirobot [40], develops all of the NLD components on a simulated robot hardware, postponing the integration and field tests with the actual hardware until later. Under the second design methodology, adopted in HARUNOBU-6 [35], a robotic travel aid to guide the visually impaired, and RHINO [9], the NLD capabilities are added after the robotic hardware is designed, developed, and deployed. Both practices increase the probability that the DMS and the robot hardware can be out of sync at run time. Under the first practice, the NLD designer must make assumptions about the robot hardware. Under the second practice, the NLD designer is forced to work with an API to the robot hardware that may not be sufficiently detailed to discover run-time problems with the hardware.

Another cause of weak self awareness may lie in the dialogue management techniques used in the DMS component. These techniques can be divided into three broad groups: state-based [33], frame-based [20], and plan-based [4]. All of them are largely deterministic. Consequently, one may question their appropriateness for describing the inherently stochastic nature of robotic hardware. Probabilistic techniques have been attempted [34], but no evaluation data are given describing their performance.

We became aware of the problem of weak self awareness during navigation experiments with our robotic guide for the blind. The cause of the problem was always the mismatch between the verbalized intent of the robot and the robot's actual actions. During several trial runs, the robot informed the navigator that it had started making a u-turn after it had already started executing the maneuver. Although the robot's message accurately described the robot's intention, it caused visible discomfort for the blind navigators. At several T-intersections the robot would tell the navigator that it was turning left (or right) and then, due to the presence of people, it would start drifting in the opposite direction before actually making a turn. When that happened, we observed that several blind navigators pulled hard on the robot's handle, sometimes driving the robot to a virtual halt. We conjecture that when a communication mismatch occurs, i.e., when the robot starts doing something other than what it said it would do, the human navigators become apprehensive and try to stop the robot.

4.4 How do robots interact with people?
As assistive technology researchers, we are interested, first and foremost, in effective and safe interaction between a disabled person and an assistive device. Effective and safe interaction must eliminate or, at the very least, considerably reduce ambiguity. So, how do robots actually interact with people? To answer this question, it is natural to look at some robots who have been made to interact with humans in various contexts. Our sample of applications cannot possibly cover the full spectrum of research on interactive social robots. For a more comprehensive survey, the reader is referred to [11, 14].

Perzanowski et al. [37] present their research at the Naval Research Laboratory (NRL) on interpreting gestures and spoken natural language commands with their mobile robot. The robot, called Coyote, interprets such commands as “turn 30 degrees left/right,” “turn to my left/right,” etc. The speech recognition system used in Coyote is the SSI PE 200. The authors conclude that numbers are difficult to understand and pose major problems for automatic speech recognition. They conclude with a hope for a more sophisticated and advanced speech recognition system to continue their work on integrating natural language and gesture.

Huttenrauch et al. [22] present a longitudinal study of one user interacting with an office delivery robot. The paper states that the speech interface is the primary interaction mode when the robot is in close proximity of the user. When the user and the robot are not colocated in the same task space, the user controls and monitors the robot via a graphical interface. However, in the actual evaluation of the system, the speech interface is not used at all. The close proximity communication is realized through a portable robot-control interface on an iPAQ Personal Digital Assistant (PDA). It is stated that the portable interface was created for the robot to be instructed in close proximity to the user without using a speech interface. No explanation is offered as to why speech was not used in the actual experiments.

Huttenrauch and Eklundh [21] present an HRI study with a service robot operating in an office environment. The robot's task is to bring coffee cups to a mobility impaired user. The robot's task is to navigate to a kitchen area and to request a coffee cup to be filled and placed on a tray mounted on top of the robot. Participants in the experiments are requested to detect the robot and listen to its spoken request for filling a cup with coffee. If a participant decides to help, he or she fills a cup from a thermos and puts the filled cup on the robot's tray. While speech-based NLD is mentioned as a possibility, speech is used only on the output. To actually communicate with the robot, the participant must locate and press a button on the robot for confirmation that the coffee is ready for delivery.

Billard [5] describes Robota, a mini-humanoid doll robot. The robot is used for educational purposes. The objective is to teach the students how to create robots with social skills. Speech is emphasized as an important social skill. The author mentions IBM's ViaVoice speech recognition system and Microsoft's SAPI. It is not clear from the article which one was actually used or if they were both used. It is stated that, to be an effective educational tool, Robota must recognize a number of key words and key phrases and generate sensible answers to queries. However, no statistics, either descriptive or inferential, are given on the usability of speech as an input or output mode.

Montemerlo et al. [34] describe the robot Pearl, which acts as a robotic guide for the elderly. In the reported experiment, 6 elderly people were guided to different locations at an assisted living institution. The researchers report problems both with speech synthesis and speech recognition, but do not give any statistical
breakdown on the types of errors and on how they affected the robot's performance or the human's perception of the robot. Gockley et al. [17] describe Valerie the Roboceptionist, a stationary robotic receptionist that resides in a small booth near the main entrance of a hall at Carnegie Mellon University and engages in NLD with people entering the hall. People can ask Valerie for office numbers and directions. While Valerie uses speech on the output, the input is restricted to the keyboard. The researchers state that the keyboard was chosen over speech, because the keyboard input is easier to control and more reliable than a typical speech recognition system.

Breazeal et al. [7] report several experiments with their Leonardo platform in which human subjects guide the robot to perform different tasks using speech and gesture. The basic tasks the subjects are required to perform with the robot are: teaching the robot the names and locations of different buttons placed in front of the robot, checking to see if the robot knows the names of the buttons, asking the robot to push the buttons, and telling the robot that the task is done. The speech understanding system uses Sphinx-4 [31] with a limited grammar to parse incoming phrases. The subjects are taught the list of allowed phrases prior to interacting with the system, which eliminates the shared vocabulary problem from the start. The list includes greetings, button labels, requests to point or press the buttons, and acknowledgments of task completions. The researchers mention speech recognition failures as one of the two common sources of error, the first being the failures of the vision system to recognize a gesture. No statistics are given on speech recognition failures, probably because the focus of the research is on the effects of non-verbal communication on the efficiency of task completion.

We can make several generic observations on NLD in robots from these investigations. First, speech as an NLD input mode is used only when the person interacting with the robot is one of the developers or when the person receives training prior to interaction. Training, however, does not guarantee against run time errors. Second, speech is used in laboratory environments where nuisance variables, e.g., human traffic and background noise, are easily controlled. Third, robots that work with people over extended periods of time in real environments either do not use NLD at all or do not use speech as an NLD input mode. Language input is solicited through GUIs or keyboards. Fourth, there is a dearth of statistically sound studies on the actual use of NLD dialogues. Presented evidence is mostly descriptive and anecdotal.

5. DECISION SUPPORT PROCEDURE
Given these observations, how can an AT researcher or practitioner evaluate the appropriateness of NLD in an assistive robot? Below we outline a decision support procedure as a series of questions that can be asked at the beginning of an assistive robotics project to determine if a given assistive robot should be dialogical. This procedure is to be viewed as a step toward a comprehensive, living document that will be of value to the AT community and will be subsequently refined and expanded through future research efforts.

- **Does the target user have any disabilities prohibiting the use of natural language?** As was stated above, certain cognitive and physical disabilities render the use of natural language, especially spoken language, practically impossible. If that is the case, some augmented communication techniques may be more appropriate than NLD.

- **Does speech misrecognition undermine the criticality of the assistive robot?** If the answer to this question is affirmative, speech is out of the question. The affirmative answer to this question led us to switch from wearable microphones to wearable keypads in our robotic guide for the blind [28].

- **Can speech misrecognition be overcome through user training or a hardware device?** This seemingly straightforward question is not easy to answer. The ability to achieve convergence through training an ASR system varies from individual to individual [23]. The most reliable way of achieving convergence is to go through the training and make the decision on the basis of obtained results. In practice, however, this frequently turns out to be an expensive option. Several hardware solutions can reduce the ambiguity caused by speech in noisy environments. One hardware solution that may reduce ambiguity in noisy environments is a push-to-speak button. To speak to the robot the user must press a button. The push-to-speak option may eliminate some ambiguity caused by continuous speech with a third party, coughs, or throat clearings. Another hardware solution is the use of the keyboard.

- **Can the target user acquire shared vocabulary?** The answer to this question has to do with the expertise level of the user and the expertise level of the vocabulary. Ideally, the expertise level at which the robot vocabulary is designed should match the expertise level of the user.

- **Is the use of NLD economically justified?** In many AT contexts, especially those that involve caregivers, e.g., assisted living homes and hospitals, there are already well established procedures and processes for doing things. Before going ahead with NLD and redesigning the existing information infrastructure, it is worth pondering if that infrastructure can be used without disruptive modification. The more disruptive the modification, the more the prospective caregiver is likely to resist.

- **How reliable is the robot hardware?** Special care is needed with proof-of-concept devices fresh out of research labs. It is not enough for the robot to perform well enough as a prototype. A good place to start is to evaluate the hardware in terms of standard task-oriented metrics for HRI [13].

- **Are speech beacons appropriate?** As we noted above, in some situations speech may not be an ideal output mode for a dialogue between the user and the assistive robot. In the absence of a sound audio perception design framework, the most reliable way of obtaining the answer is pilot experiments.

6. THREE CONJECTURES
In this section, we will attempt to generalize our discussion and speculate when NLD might be appropriate for HRI in general. We
subscribe to the view that the ultimate objective of robotics is the production of effective servants, not alternative life forms [18]. The robot must do what its user tells it to do when the user tells it to do so, not when the robot itself considers it fit to do so. Viewed under this bias, the relevant question is whether NLD is the most effective means of telling the robot what to do and receiving feedback from the robot.

To speculate on this question, let us postulate three real-valued variables: $A$, $I$, and $P$. $A$ measures how autonomous the robot is, $I$ measures how much the human operator must intervene in the actual operation of the robot, and $P$ measures how much potential exists for NLD. Let us further agree that all variables are in the closed range from 0 to 1. Let us now make and discuss a few conjectures about $P$.

**Conjecture 1: As $A$ goes to 1, $P$ goes to 0.**

This conjecture states that the closer the robot is to full autonomy, the smaller the potential for NLD. In the extreme case of a factory assembly robot, there is no need for NLD. Generally speaking, the more the robot is capable of making and executing its own decisions autonomously, the less we need to interact with it through NLD. Full autonomy is currently achievable only in restricted environments. A fully autonomous robot, by definition, does routine, repetitive work that a human is not interested in doing.

**Conjecture 2: As $I$ goes to 1, $P$ goes to 0.**

This conjecture states that as the level of human intervention increases, the potential for NLD decreases. Language is a slow medium, because producing and interpreting utterances takes time. A larger amount of human intervention implies that the operator is involved in the decision making process of the robot. This, in turn, implies that the operator is likely to have a formal model of what the robot is doing and why. Formal models are expressed in formal languages for which natural language is not appropriate due to its ambiguity and vagueness. There is a reason why the Mars Rover [30] did not have an NLD component: it is always faster for a knowledgeable operator to interact with the underlying computational model directly rather than through a slow intermediary such as natural language.

**Conjecture 3: $P > 0$, when $A \in [a_1, 1−\delta_1]$, where $0 < \delta_1 < 1$, $0 < a_1 < 1−\delta_1$, and $I \in [a_2, 1−\delta_2]$, where $0 < \delta_2 < 1$, $0 < a_2 < 1−\delta_2$.**

This conjecture states that the potential for NLD exists when the robot is partially autonomous and requires some human intervention. Obviously, $A$ and $I$ are inversely related. But how high should $A$ be for NLD to be justified? We conjecture that $A$ should be at 0.5 or above. In other words, the robot should be physically competent. Unfortunately, that does not seem sufficient. The robot's physical competence should also be cognitively interesting for the human. It is not all that exciting to have a dialogue with a robot that can only make turns and change rotational and translational velocities. As argued by Brooks [8], language may have been a late arrival on the road of evolution. Living forms had been working for millennia on their basic physical and perceptual skills before language appeared on the scene. Even if one does not accept the evolution argument, one is likely to agree that for language to be justified our robots must first become physically and cognitively interesting to us.

7. CONCLUSIONS

We examined the claim that NLD is a promising HRI mode in the context of assistive robotics. We defined assistive robots in terms of an existing HRI taxonomy. Based on the results of our analysis, we outlined a preliminary decision support procedure for AT researchers and practitioners who need to evaluate the appropriateness of NLD in assistive robots. Finally, we offered several conjectures about when NLD may be appropriate as an HRI mode in general.

8. ACKNOWLEDGMENTS

The author would like to acknowledge that this research has been supported, in part, through NSF CAREER grant (IIS-0346880) and two Community University Research Initiative (CURI) grants (CURI-04 and CURI-05) from the State of Utah. The author would like to thank the participants of the 2005 AAAI Spring Workshop on Dialogical Robots at Stanford University for their comments on his presentation “Should Rehabilitation Robots be Dialogical?” The present paper has benefited a great deal from the workshop discussions. The author is grateful to four anonymous reviewers for their comments.

9. REFERENCES


[37] Perzanowski, D., Schultz, A., and Adams, W. Integrating natural language and gesture in a robotics domain. In IEEE International Symposium on Computational Intelligence in


