From Commands to Goal-based Dialogs: A Roadmap to Achieve Natural Language Interaction in RoboCup@Home

Mauricio Matamoros[0000−0002−4054−9079], Karin Harbusch, and Dietrich Paulus

Active Vision Group (AGAS), University of Koblenz-Landau.
Universitätsstr. 1, 56070 Koblenz, Germany

Abstract On the one hand, speech is a key aspect to people's communication. On the other, it is widely acknowledged that language proficiency is related to intelligence. Therefore, intelligent robots should be able to understand, at least, people's orders within their application domain. These insights are not new in RoboCup@Home, but we lack of a long-term plan to evaluate this approach.

In this paper we conduct a brief review of the achievements on automated speech recognition and natural language understanding in RoboCup@Home. Furthermore, we discuss main challenges to tackle in spoken human-robot interaction within the scope of this competition. Finally, we contribute by presenting a pipelined road map to engender research in the area of natural language understanding applied to domestic service robotics.

Keywords: Robotic competitions · Natural Language Understanding · Artificial intelligence and robotics.

1 Introduction

From its foundation, RoboCup@Home has stressed the importance of natural interaction between humans and robots. With the target perspective that intelligent robots can understand people's orders that fall within their application domain. Thus, Human-Robot Interaction (HRI) has always been pursued in the competition.

However, a detailed evaluation of natural language interactions is not easy. In favor of many other functions in early stages of the robot development, simple straightforward commands often served the purpose of interaction. This, along with a) the noisy competition environments, b) the biased, non-native speaker operators; and c) the use of command generators to instruct robots have impeded the definition of fine-grained evaluation measures RoboCup@Home for natural language understanding in the domain of RoboCup@Home.

In order to achieve this goal, we first analyzed the progression of the league in Automatic Speech Recognition (ASR) and Natural-Language Understanding (NLU) in the last nine years. Our proposal is based on claims made by the participating teams in their Team Description Papers (TDPs), relevant publications,

rulebooks, multimedia material available on-line, and our cumulative experience as participants and referees in RoboCup@Home since 2009.

In this paper we present a road map to pave the way towards a completely natural interaction between humans and robots. This ultimate goal is achieved by defining milestones that promote the use of natural-language interaction. We underpin our strategy by getting the general public involved in the creation of a large annotated dataset of untrained and unbiased operators, inexistent to the extent of our knowledge. Here, the competition plays a fundamental role, since RoboCup@Home sets the perfect scenario to involve the audience in data production for scientific use. We believe this data will foster research in both, Natural-Language Processing (NLP) and HRI.

The paper is organized as follows. In Section 2, we briefly introduce the RoboCup@Home league. In turn, we present a summary of the last nine years of the competition, and give an overview of the strategies used by the teams to solve the tests. In Section 3 we summarize the main unresolved challenges in huma-robot interaction. In Section 4 we discuss a roadmap and its milestones to deal with the key problems. Finally, in Section 5 we sum up and draw conclusions.

2   Speech Recognition and Natural Language Understanding in RoboCup@Home

RoboCup@Home aims at developing service and assistive robot technology by evaluating a robot's performance with a series of **tests** in an unstandardized and realistic scenario. Here, new approaches are tested in a competitive setup where robots have to solve a set of common household **tasks** upon request. Consequently, the competition influences research, even sometimes directing it. Such is the case of ASR and NLU, since they are fundamental to HRI.

In a **test**, a task is divided in a sequential set of subgoals or phases, being necessary to fulfill one goal to advance to the next phase [8]. This approach keeps difficulty reasonably low for newcomers, while still striving for high top level performance [25]. Usually, in the beginning, the **operator** gives a spoken command to the robot. A **command** is typically a short imperative sentence containing all required information to execute a predefined task. Normally, the operator and the **referee** take account of the robot's processing, whereas the scoring is executed by the Technical Committee (TC), which is also in charge of design tests and review rules. Detailed descriptions can be found in [27,8].

In RoboCup@Home, ASR and NLU are closely related. Spoken commands have always been the preferred way to operate a robot and, since robots need to confirm they have understood the command, both abilities are commonly scored together.

The evolution process of Automatic Speech Recognition and Natural-Language Understanding in RoboCup@Home is addressed as follows. In Section 2.1, we summarize the last nine years of competition, focusing on the main challenges the teams had to face. In Section 2.2, we present an overview on strategies used to solve the tests.

2.1 Historical Overview of HRI testing

Our review starts in 2009. We have chosen this year because the rulebook of 2009 is the oldest available on the RoboCup@Home website[1].

For the competition in **Graz 2009** the guidelines for operating a robot were loose. Most interactions were hardwired, based on each team's preconceptions of what a natural HRI could be. In addition, the use of headsets and wireless microphones was common, and some tests rewarded the use of gestures over speech.

Istanbul 2010 came with important changes: a) scoring was made explicit, b) score sheets were included in the rulebook, and c) interaction guidelines were included as part of the tasks in each test. For instance, a robot could score for catching the name of a person or room (typically, only the name was given).

But the cornerstone of 2010 was the inclusion of the *General Purpose Service Robot* (GPSR) test. In this test the operator could command the robot to perform any task from any of other test (including those of former years). Moreover, robots had to deal with long sentences and incomplete information for the first time; a big step in NLU and HRI. Notwithstanding, all interactions in GPSR followed the patterns used in the other tests.

For **Singapore 2011** most interaction guidelines had been removed and a **command generator**[2] was developed by the TC for GPSR (teams only had access to a limited version).

No significant changes in NLP or spoken HRI were introduced in the next three years. Nevertheless, the TC noticed a sustained performance decrease in command retrieval.

To help teams, in **João Pessoa 2014** the TC introduced a way to bypass speech recognition in order to solve the task: the ***Continue* rule**. However, only few teams took advantage of it.

For **Hefei 2015** the use of QR codes to bypass ASR was made compulsory. Moreover, the data recording feature allowed teams to get a scoring bonus and contributing with the league by providing all data acquired by the robot during a test. In addition, a new command generator was open sourced. Notwithstanding, even having access to the verbatim output of the GPSR command generator via the QR Code, many robots remained unresponsive.

In consequence, **Leipzig 2016**, the TC decided to provide open access to the command generator and the generation grammars about one month before the competition. This was a crucial decision due to its direct impact on natural HRI, as explained in Section 3. Cloud Computing was another minor but important change. Prior to this year, the availability of an Internet connection through the arena's wireless network wasn't granted. However, in combination with the low reliability of the network, discouraged teams from exploring solutions based on cloud services.

By **Nagoya 2017**, the previous two years of tests focused on benchmarking had paid off (See Table 2). In addition to an increase in performance in ASR,

---

[1] http://www.robocupathome.org

[2] Available on-line: http://komeisugiura.jp/software/2010_GeneralPurposeTest.tgz

the relevance of command interpretation and NLU grew with the inclusion of the Standard Platform Leagues (SPL). The out-of-the-box-ready robots allowed teams to focus on high-level problems such as command interpretation, NLU, and task planning and reasoning. This year was the first time in which robots had to answer questions about their environment and previously executed tasks. Moreover, in the Tour-Guide test they also had to attract people outside the competition area, introduce themselves, and answer people's questions without the help of any grammar.

2.2   Adopted Strategies and Software Solutions for HRI

Either in face-to-face communication or remotely like by phone, radio, or TV, the hearer decodes the produced sounds of the speaker, trying to match them with the best interpretation given the current context. Similarly, spoken Human-Robot Interaction requires the extraction and analysis of the language elements of the uttered sentence.

Figure 1: One of the most common processing chains for spoken commands

Although possible, dealing with audio signals can be extremely difficult in high levels of abstraction [16,3]. Therefore, the most broadly adopted solution consists in using an Automatic Speech Recognition (ASR) engine to get a text-transcript for further processing as depicted in Figure 1. Typically, **cardioid microphones** are used as main audio input for the ASR engine to cope with noise. Afterwards, the ASR engine output is preprocessed by some sort of [natural] language processor so the task planner can select the most adequate behavior.

Other than the microphone itself, filters are normally absent and the filtering task is

Table 1: Most used ASR software in 2017

Usage	ASR Engine Name
25.81%	CMU [Pocket] Sphinx
16.13%	Google Speech Api
16.13%	Microsoft Speech Api
12.90%	Julius
9.68%	Nuance VoCon/DNS
6.45%	Rospeex
6.45%	Intel RealSense SDK
3.23%	Amazon AWS / Alexa
3.23%	Kaldi
3.23%	iFlyTek
9.68%	Unreported

delegated to the ASR engine itself [4]. Although several noise-reduction filters have been tested in the competitions, we couldn't find any reported successful solution other than **HARK**[3] [14,2,7]. However, HARK is used mostly for sound-source localization and separation.

---

[3] HARK (**H**onda Research Institute Japan **A**udition for **R**obots with **K**yoto University) is an open-source robot audition software that includes modules for ASR and sound-source localization and sound separation. Source: https://www.hark.jp/

Continuing with the pipeline of Figure 1, the most commonly adopted software solutions include Loquendo (now Nuance) ASR [22], and the Microsoft Speech API [17], being most popular **CMU Sphinx** [19,23] as Table 1 shows. However, in 2017, due to the limited computing power of the robots in the SPL, most teams used cloud services (mainly the Google speech API [14,26]) or relied on the built-in ASR system [12,24] which, as the rightmost two columns of Table 2 show, were not as good as other solutions.

Table 2: Top-10 scores in ASR-related tests
(final rank in parentheses)

Rank	2015	2016	2017		
			OPL	DSPL	SSPL
1	86.7% (1)	86.7% (3)	96.7% (1)	67.3% (1)	71.5% (1)
2	83.3% (2)	83.3% (6)	83.3% (3)	45.3% (5)	52.1% (2)
3	70.0% (8)	72.7% (13)	73.3% (5)	37.3% (3)	36.4% (4)
4	70.0% (10)	60.7% (2)	70.0% (4)	12.0% (4)	30.3% (5)
5	66.7% (6)	60.7% (5)	66.7% (2)	3.3% (7)	29.1% (6)
Top5 $\overline{x}$	75.3%	72.8%	78.0%	33.0%	43.9%
Top5 σ	9.0%	12.2%	12.2%	25.8%	18.0%

Despite the remarkable advances achieved in Natural-Language Processing and task planning in recent years, RoboCup@Home has taken little advantage of it. Some reasons include a) the sequential nature of the tests, b) the simplicity of the tasks (HRI-wise), c) the computational power available in the robot, d) the lack of awareness due to sensors' limitations, and e) the need of recognizing only a few words. In consequence, most approaches for NLP relied in **keyword spotting** or pattern matching to trigger the execution of a **state machine** [20,15]. At least in the beginning, this strategy seemed to be faster and more robust than its more advanced counterparts, albeit much simpler.

Despite this, robust A.I. solutions have always been in play. As of 2013, it was common to find the task planner and the natural language processor fused in the same module, which doesn't seem to be the case anymore. Common strategies included a) the use of rules of inference for both, sentence parsing and task planning [11]; b) probabilistic parsers [5], c) semantic networks [17], and d) regular expressions [15] to name some.

Although keyword spotting, **pattern matching**, and state machines are currently used solutions (specially in simple tests), we spotted more powerful and avant-grade approaches. Whilst in 2017 only 39% of the teams didn't mentioned neither NLP nor NLU approaches in their Team Description Paper (TDP), in all other reports the task planner and the NLP are separated. Among the approaches for processing language, we found a) Probabilistic Semantic Parsers [6]; b) Multimodal Residual Deep Neural Networks [10], c) ontology-based parsers over inference engines [18], and d) probabilistic parsers for syntax-tree extraction along with lambda calculus for semantic parsing [9]. The **Stanford Parser** [21] is the most broadly adopted solution for POS-tagging and syntactic tree extraction, and **LU4R** [1], a Spoken Language Understanding Chain for HRI developed in

La Sapienza [13] by participants of RoboCup@Home which is used by several teams.

These newer solutions have increased the robots' performance in the latter versions of GPSR when some ambiguity was added. Nevertheless, the league does not acquire these newer approaches. Most robots are still using grammar-based ASR engines, which limit the input of the NLP unit. Moreover, so far no test requires the resolution of ellipsis or anaphora. Finally, another major inconvenience, is the use of command generators instead of natural interactions. How to develop HRI tests which focus on these problems is discussed in Section 3.

3   Challenges

As stated in previous sections, several setbacks have been holding back the league's advances in natural language dialogs with the robot. Here we exemplify challenges and come up with guidelines for future research when addressing tougher scenarios.

3.1   Noise

One of the most problematic aspects (and to which most attention has been paid) is the noise in the competition environment. Having a separated hall as in 2012 didn't help much. The **ambient noise** produced by over two hundred people greatly exceeds the noise levels of an average apartment.

In the past, several solutions have been proposed to this problem. Having a separate hall, arenas with tall walls, and transform the arena in a sound-proof closed area with glass walls are recurrent examples. However, service robots will also operate in noisy environments such as airports and shopping malls. Thus, we think it is best to deal with this issue in an early stage. For this reason, noise is addressed in the roadmap presented in Section 4.

3.2   Operators

An aspect that has characterized RoboCup@Home is that the league has always been robot-friendly. People volunteering as operators in RoboCup@Home are often patient and prone to follow the robot's instructions, unconsciously trying to help the robot to succeed (e.g. repeat a given command louder when the robot seems unresponsive). In addition, almost all the operators are specialists, or at least familiar with service robotics. In other words, all testing has been performed by (unconsciously) **biased operators**, making it easier for the robots to accomplish their tasks. However, trying to give a positive impression to the audience, has the disadvantage of voiding the purpose of RoboCup@Home of providing real-case scenarios for testing.

Another important closely related aspect addresses the demographics of the league's participants. An international community will offer **diverse accents** to the robots and styles (which may not be correct). But this diversity comes with a

price. With few exceptions, operators having English as second language lack the richness and verbal fluency of a native speaker. Same hold for gender and age.

Therefore bias, variety of speech and diversity of accents, lexicon, and styles, are also addressed in the roadmap of Section 4.

3.3   Generators

Although the rulebooks never provided interaction templates, suggested guidelines and the sentences produced by the command generators codify biased HRI in a similar way. Besides, officializing the release of the command generator and its **grammars** had the pernicious effect of replacing natural language with a simplistic one, i.e. artificializing interactions. Moreover, the person in charge of the generator (one of the authors of this paper), although proficient, is not a native speaker of the official competition language (US-English). In summary, despite stimulating immediate positive results and supplying the teams with a powerful tool, the careless use of generators might delve into a ballast in the long term. For this reason, the use of command generators is revised in the roadmap of Section 4.

In order to overcome the problems discussed in this section, the proposed solution should force robots to deal with a vast diversity of unbiased operators speaking freely and under reasonable conditions of noise, while keeping tests fair and scientifically meaningful. This ultimate goal is analyzed and split into small steps which are presented in Section 4.

4   Solution strategy and roadmap

Ideally, the robots will lead a natural-language dialogue in real-environment conditions. This is the ultimate goal in service robotics regarding NLU. Considering the challenges presented in Section 3 we propose a series of specific tests and changes to the existing ones.

In the previous section we identified the biased operators in combination with restricted GPSR as a main obstacle for the stagnation in elaborate HRI. In order to set up a rich corpus of human-robot dialogs, we opt for recruiting untrained English native speakers as **untrained operators**, e.g. selected from the audience. The **unbiased operators** must have little to no previous experience in RoboCup@Home (or preferably in robotics), and be allowed to interact freely with the robot. With these changes, we are solving biasing, variety and diversity of speech, and tackling the artificiality introduced with the command generators.

Needless to say, the proposed changes towards free natural-language dialogs with the robot have to be applied gradually. Therefore, in this section we present a roadmap of three phases that takes as axis the *General Purpose Service Robot* test before changes are propagated and adapted for other tests. Each phase sets a constraint that will rule over the upcoming featured milestones. Milestones present **small increments** in the difficulty of the NLU task, presumably achievable with the data collected in former years. Once solved in GPSR, the constraints

of a milestone would become standard practices in other tests. In addition, milestones are planned to respect the two-years-limit established by the founders of the league [8]. Furthermore, the proposed roadmap has the advantage of being adaptable. If sufficient performance hasn't been achieved, the milestones can be shifted forward in time.

In order not to overtax the teams, we propose the following 3-step pipeline to implement the changes by small increments towards free natural-language dialogs with the robots. First, teams test and benchmark with recordings addressing the features of the milestones newly defined; then, those features are tested in GPSR; and finally, propagated to all other pertinent tests.

Hence, the strategy considers also the inclusion of a *Natural-Language Understanding and Action Planning* benchmark in Stage I with a small contribution to the overall score. In this benchmark, a team receives a set of **sound files**, each one containing a task-execution request from an unbiased operator of the same kind that would be given in GPSR. The A.I. of the robot needs to transcribe and analyze these recordings, extracting either a goal, a plan with a set of actions to carry out, or a set of questions or statements to continue the interaction. The score should consider not only the analytical quality of producing transcripts (where applicable penalizing hard-wired constructions in favor of generalized NLP-rules), but also how far the robot went in its planning, and if it was following the right direction. In this way, milestones could be tested one or two years before being implemented in GPSR and propagated to other tests, giving time to teams to prepare. Furthermore, by giving the same recordings to all teams, **fairness** and replicability are addressed, while noise can be tackled by superposing ambient **noise** to the recordings of the operators. It will be up to the TC to decide on the nature and intensity of the noise regarding the league's advances in the subject.

In addition, to support the league and foster research, the proposed strategy exploits the **Data Recording** feature, incorporated in 2015 to the competition. This is deemed as fundamental for the success of our proposal. We think all interactions between the operator and the robot should be recorded and distributed under **Open Access** license as soon as a transcription is available. This includes all benchmarking sets from previous years. In this way, the league supports both, experienced and new competitors. Beyond this obvious assistance the entire scientific community gains relevance. In early stages, teams providing recordings of all speech-based interactions during a test might receive a bonus proportional to the score obtained, as well as an additional bonus or a certificate (as decided by the TC) for annotating the provided data. Later on, such policy should be made compulsory for the roadmap to work with full efficiency, automating the collection process if possible.

Before summing up by means of presenting the roadmap, we have to state some minor clarifications: 1) In all phases, operators with no background in robotics are selected from the audience by the TC. 2) Unless the test specifies otherwise, chosen operators shall be fluent English speakers. 3) Commands, goals, and tasks are generated before the test. A referee must explain all pertinent information to the operators before spoken interactions are recorded and noise

overlapped. 4) The TC randomly assigns to each robot a set of generated tasks from the pool. The robot will listen to the recording, but the operator must be present for further (unexpected) interactions.

Roadmap phase 1: From commands to goal formulation

The referee gives a specific goal to the operator along with all pertinent information regarding the desired task. For this purpose, a random command generator can be used. When required, the operator can practice with the referee. Referees assist operators in unexpected situations.

The intensity of overlapped noise should increase gradually, starting from relatively quiet environments (country house, city apartment, office, etc) towards moderately noisy ones (busy office, or a restaurant with background music).

Yearly progress
2019 The operator reads the generated command.
        Note: This vanilla milestone introduces no change to allow the benchmarking of the next milestone in the pipeline.
2020 The operator tries to memorize the command, repeating it to the robot afterwards. Slight variations are expected.
2022 The referee explains the task to the operator, who has to command the robot using their own words (e.g. rephrasing).
2024 The operator requires the robot to accomplish a task, although not necessarily in an imperative way, as if suggesting.
2028 The operator tries to explain the goal to the robot, not specifying what to do, but the expected final result.

While no big changes are expected during the first four years in NLP, we foresee the inclusion of more robust filters and the use of less constrained grammars. However, by the third milestone (2022), operators might unintentionally neglect information, so robots will need to reconcile information as it arrives, making **questions** as needed. Furthermore, people normally make a very efficient use of language, so references are very common. Therefore, we expect the exploration of **reference resolution** such as ellipsis and anaphora by this year. Finally, the latter milestones will keep the pace, slightly moving the focus to task planning while addressing new types of sentences.

Roadmap phase 2: Towards dialog-based interaction

The referee gives to the operator a set of examples of what robots can do in a given domain. The operator must propose a similar task for the robot which has to be approved by the referee. The robot may not know how to accomplish the task, in which case the procedure is explained by the operator.

The intensity of overlapped noise ranges from medium to high, with sudden bursts of loud recorded human voices like in shopping malls, grocery stores, and airports are good examples.

Yearly progress
2032 After receiving a set of examples, the operator elicits a similar behavior to the robot. The procedure can be explained in detail step by step.

2036 After receiving a topic and a set of examples, the operator requests
      something of the same difficulty. Sub-goals are detailed to the robot, but
      not the individual commands. The operator corrects the robot's plan.

In these phases, we continuously add new elements to NLU, integrating it deeper with **planning**, while, at the same time, we are collecting new applications from users (the potential market). Plan-learning using natural language requires the integration of **cardinality**, as well as spatial and temporal relationships (e.g. the operator may request *make me a sandwich*, explaining later on the steps like *grab a slice of bread, spread mayo...*).

In contrast to only presenting an initiated plan to the operator, the tasks stimulate that the operator and the robot have to enter in a **dialog** towards an unconstrained natural interaction.

Roadmap Goal: Reaching unconstrained Interaction

In the last phase, the operators only get explained the State-of-the-Art, i.e. the limits of what the robots can achieve. They can request anything crossing their minds within those limits the way they want. The robot may not know how to accomplish the task, in which case the procedure is freely explained by the operator in a dialog.

Yearly progress

2040 The operator explains to the robot how to perform an entirely new task
      and what the results should be.
2044 The operator requires anything from the robot within the state-of-the-art.
      All planning is left to the robot.

We believe the presented roadmap helps the RoboCup@Home community to push the boundaries of research in HRI.

The first steps will force teams to look for alternatives to grammar-based ASR engines, or at least a less restricted ones. Besides, the continuous analysis of audio signals to separate the operators' voice from noise is also addressed, although dealing with noise can be left as an option for daring teams.

Regarding NLU, the first steps will take current approaches to the limit. Unbiased operators will gradually expose robots to the richness of **free speech**, while the pipeline will grant time to prepare.

Later on, moving from imperative sentences (e.g. *clean the bedroom*) to goal-driven sentences of any kind (e.g. *I need the room clean for tonight, is the breakfast ready?*) will not only have the league working with semantics and pragmatics, but also might foster research in action planning.

Especially, in the second phase, more complex elements of language analysis are incorporated, paving the road towards a dialog-based interaction. On that basis, we stipulate conversational robots in RoboCup@Home. This claim is substantiated in the last phase by widening the domain and removing all constrains from the operators, in order to deal with real-world situations.

Nevertheless the roadmap provides the necessary flexibility for the worst-case scenario where direct instructions can be given to the robots.

5   Conclusions

In this paper we provided a historical overview of testing natural-language interactions with robots over the last nine years of RoboCup@Home. We outlined the state-of-the-art Automatic Speech Recognition and Natural-Language Processing. We quantified recent strategies and software solutions adopted by teams to overcome the trials set in the competition.

In these observations we identified a set of challenges that haven't been tackled. Critical components that hamper free dialogs with the robot are posed by the command generator along with the subconscious unintentional bias of the operators. We inspected how individual test features prevent the league from dealing with free natural language. Based on this study, we propose a strategy and a roadmap that formulate key features to implement the in the individual NLP components of the robot (see Figure 1) to conquer those challenges. Many test details strongly depend on the advancement of ASR engines, neglects non-verbal communication, so there should be an entangled roadmap for these features.

We hope this roadmap will serve as initiative to promote long-term planning in RoboCup@Home. In particular our contribution elaborates on stepstones to elicit thorough NLP to empower robots with strong natural-language skills.

References

1. Bastianelli, E., Croce, D., Vanzo, A., Basili, R., Nardi, D.: A discriminative approach to grounded spoken language understanding in interactive robotics. In: IJCAI. pp. 2747–2753 (2016)
2. Demura, K., Demura, K., Nagashima, K., Enomoto, K., Yamakawa, T., Iwasaki, R., Mashimo, S.: Happy mini 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)
3. Dominey, P.F.: Learning grammatical constructions from narrated video events for human–robot interaction. In: Proceedings IEEE humanoid robotics conference, Karlsruhe, Germany (2003)
4. Doostdar, M., Schiffer, S., Lakemeyer, G.: A robust speech recognition system for service-robotics applications. In: Iocchi, L., Matsubara, H., Weitzenfeld, A., Zhou, C. (eds.) RoboCup 2008: Robot Soccer World Cup XII. pp. 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
5. Gaisser, F., Aswin Chandarr, A., Rudinac, M., Bruinink, M., Pons, S., Rueda, M.B., Lung, G.L., Wisse, M., Jonker, P.: Delft robotics robocup@home 2013 team description paper. Proceedings RoboCup competition (2013)
6. Hart, J.W., Stone, P., Thomaz, A., Niekum, S.: Ut austin villa robocup@home domestic standard platform league team description paper. RoboCup @Home 2017 Team Description Papers (2017)
7. Hori, S., Ishida, Y., Kiyama, Y., Tanaka, Y., Kuroda, Y., Hisano, M., Imamura, Y., Himaki, T., Yoshimoto, Y., Aratani, Y., Hashimoto, K., Iwamoto, G., Morie, T., Tamukoh, H.: Hibikino-musashi@home 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)
8. Iocchi, L., Holz, D., Ruiz-del Solar, J., Sugiura, K., Van Der Zant, T.: Robocup@home: Analysis and results of evolving competitions for domestic and service robots. Artificial Intelligence **229**, 258–281 (2015)

9. Lee, B.J., Choi, J.Y., Lee, C.Y., Park, K.W., Choi, S., Baek, C., Zhang, B.T.: 2017 aupair team description paper. RoboCup @Home 2017 Team Description Papers (2017)

10. Liu, J., Zhang, Z., Tang, B., Chen, X.: Wrighteagle@home 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)

11. Llarena, A., Boldt, J.F., Steinke, N.S., Engelmeyer, H., Rojas, R.: Berlinunited@ home 2013 team description paper. Proceedings RoboCup competition (2013)

12. Martínez, L., Muñoz, R., Olave, G., Pais, G., Hernan, G., Gomez, D., Garrido, L., Campanini, D., Orellana, P., Loncomilla, P., Ruiz-del Solar, J.: Uchile homebreakers 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)

13. M.T. Lázaro, Iocchi, L., Nardi, D., Hanheide, M., Fentanes, J.P.: Spqrel 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)

14. Oishi, S., Miura, J., Koide, K., Demura, M., Kohari, Y., Une, S., Villamar-Gomez, L., Kato, T., Kojima, M., Morohashi, K.: Aisl-tut @home league 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)

15. Pineda, L., Meza, I., Fuentes, G., Rascón, C., Peña, M., Ortega, H., Reyes-Castillo, M., Salinas, L., Ortega, J., Rodríguez-García, A., et al.: The golem team, robocup@home 2013 (2013)

16. Roy, D.K., Pentland, A.P.: Learning words from sights and sounds: A computational model. Cognitive science **26**(1), 113–146 (2002)

17. Savage, J., Negrete, M., Matamoros, M., Cruz, J., Contreras, L., Pacheco, A., Figueroa, I., Márquez, J.: Pumas @home 2013 team description paper (2013)

18. Savage, J., Negrete, M., Cruz, J., Marquez, J., Martell, R., Cruz, J., Vazquez, E., Pano, M., Cruz, J., Silva, E., Estrada, H., Arce, H., Matamoros, M., Garzon, A., Fuentes, O.: Pumas@home 2017 team description paper. RoboCup @Home 2017 Team Description Papers (2017)

19. Schiffer, S., Niemüller, T., Doostdar, M., Lakemeyer, G.: Allemaniacs @home 2009 team description. Proceedings CD RoboCup (2009)

20. Seib, V., Manthe, S., Memmesheimer, R., Polster, F., Paulus, D.: Team homer@unikoblenz—approaches and contributions to the robocup@home competition. In: Robot Soccer World Cup. pp. 83–94. Springer (2015)

21. Stanford: Corenlp (2011), http://nlp.stanford.edu:8080/corenlp/

22. Stückler, J., Dröschel, D., Gräve, K., Holz, D., Schreiber, M., Behnke, S.: Nimbro @home 2010 team description (2010)

23. Wachsmuth, S., Lier, F., Meyer zu Borgsen, S., Kummert, J., Lach, L., Sixt, D.: Tobi - team of bielefeld a human-robot interaction system for robocup@home 2017. RoboCup @Home 2017 Team Description Papers (2017)

24. Williams, M.A., Pfeiffer, S., Vitale, J., Tonkin, M., Ojha, S., Billingsley, R., Alam, S., Kang, L., Gudi, S.L.K.C., Clark, J., Wang, X., Johnston, B.: Uts unleashed! for robocup 2017 @home spl. RoboCup @Home 2017 Team Description Papers (2017)

25. Wisspeintner, T., Van Der Zant, T., Iocchi, L., Schiffer, S.: Robocup@home: Scientific competition and benchmarking for domestic service robots. Interaction Studies **10**(3), 392–426 (2009)

26. Yaguchi, H., Tran, B., Takeda, M., Kochigami, K., Li, Z., Sasabuchi, K., Furuta, Y., Nagahama, K., Okada, K., Inaba, M.: Jsk@home: Team description paper for robocup@home 2017. RoboCup @Home 2017 Team Description Papers (2017)

27. van der Zant, T., Wisspeintner, T.: Robocup@home: Creating and benchmarking tomorrows service robot applications. In: Robotic Soccer. InTech (2007)