ingenieur
wissenschaften
htw saar

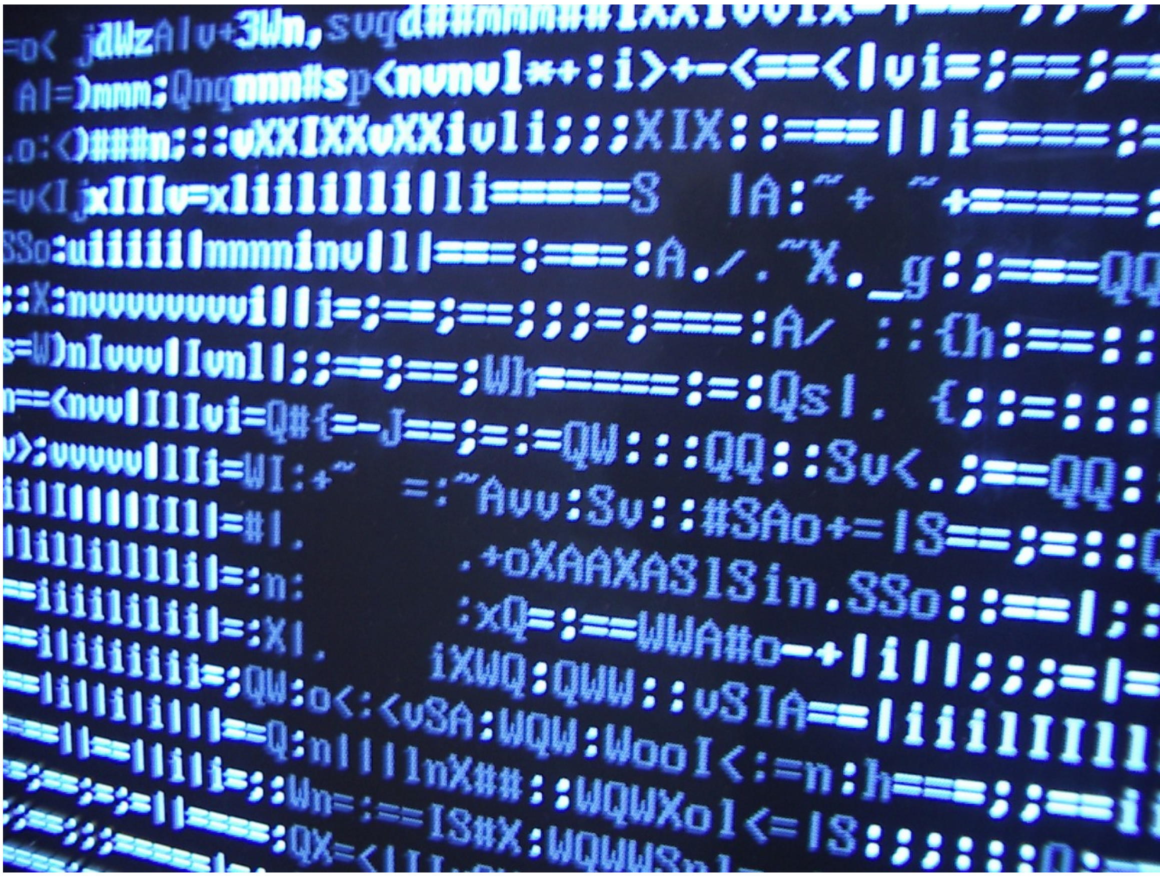
Hochschule für
Technik und Wirtschaft
des Saarlandes
University of
Applied Sciences

STL SYSTEMTECHNIKLABOR

A Natural Language Approach to Human-Robot Interaction

Marek Kohn

Technical Report – STL-TR-2017-01 – ISSN 2364-7167



stl.htwsaar.de

Technische Berichte des Systemtechniklabors (STL) der htw saar
Technical Reports of the System Technology Lab (STL) at htw saar
ISSN 2364-7167

Marek Kohn: A Natural Language Approach to Human-Robot Interaction
Technical report id: STL-TR-2017-01

First published: July 2017
Last revision: July 2017
Internal review: André Miede

For the most recent version of this report see: https://stl.htwsaar.de/

Title image source: Katia Grimmer-Laversanne, http://de.freeimages.com/photo/digital-dream-2-1456675

htw saar – Hochschule für Technik und Wirtschaft des Saarlandes (University of Applied Sciences)
Fakultät für Ingenieurwissenschaften (School of Engineering)
STL – Systemtechniklabor (System Technology Lab)
Prof. Dr.-Ing. André Miede (andre.miede@htwsaar.de)
Goebenstraße 40
66117 Saarbrücken, Germany
https://stl.htwsaar.de

A Natural Language approach to Human-Robot Interaction

Marek Kohn
htw saar – Hochschule für Technik und Wirtschaft des Saarlandes
Seminar "Computer Science and Society"
Sommersemester 2017

***Abstract*—With HRI experiencing a huge increase in interest over the last years, we have to think of ways how to improve our coexistence with robots. Since service robots are now applied in areas, where HRI has never played a major role before, new paradigms need to be developed to establish a more natural interaction with robots. Despite the fact that natural language is considered to be one of the best and most natural ways to communicate with robots, there are certain situations where a merely verbal interaction might fail. However, making use of the possibilities that a robot provides, it might be able to supplement a verbal conversation with nonverbal modalities. In this paper we will examine how and to which extent a dialog system in HRI can be enhanced by applying a multimodal system, providing additional information and a sense for semantic context.**

I. Introduction

The idea of robots supporting humans in their everyday life has a long-standing historic background. One of the earliest concepts was designed by Hero of Alexandria, 85 A.D., who created an operative machine that is capable of automatically pouring wine for party guests [1]. Ever since then, several attempts have been made in order to imitate human abilities with the help of dedicated machines, which we usually refer to as robots.

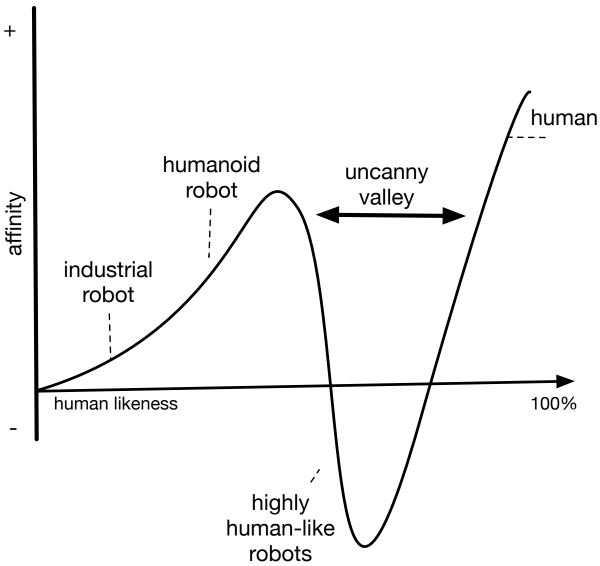


Figure 1. The *Uncanny Valley* describes the phenomenon that a higher level of human likeness does not automatically result in greater affinity in society. Referring to [9].

In general a robot can be defined as an embodied artificial intelligence, that is capable of sensing its environment and automatically makes useful work [2]. Especially from an ethical point of view, an essential demand to robotic systems is that they should always act non-destructive and to the advantage of the human race. (cf. *Three laws of robotics*) [3]:

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Traditional applications of robots are often situated in industrial contexts and thus require very little human interaction. Hence industrial robots are usually controlled remotely by human supervisors with the aim to accomplish results-oriented tasks. This often requires a technical expertise on the supervisor's side. More recently robots have expanded their usage in industry into a collaboration with humans in civil and domestic spaces [4]. Characterised as service robots, most of them are designed for convenient purposes – think of vacuum cleaning robots – but can also act as social companions and service providers [5]. Those, typically having a human-like appearance, are commonly referred to as *Humanoids* or *Androids*. As they are designed to engender confidence, Humanoids are often applied in social services such as caregiving, healthcare, elder care [6] or even security [7]. Thus it is inevitable that society accepts these robots as peers. However, studies have proven that the affinity towards entities only increases with their human likeness until a specific point is reached. Thenceforward the majority of people rather rejects entities that are to close to their own appearance. Simply put, humans tend to favour robots with "cute", human-oriented looks over robots that eerily imitate human attributes [8]. Figure 1 shows the phenomenon which is called the *Uncanny Valley* and was first described by Mori in 1970 [9].

With regard to overcome the Uncanny Valley, there are several approaches that concentrate on improving the interaction between humans and robots [10]. Goodrich (2007) comprises them within a dedicated research area, commonly referred to

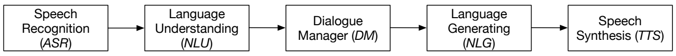


Figure 2. Basic architecture of a conversational interface

as *Human-Robot Interaction* (HRI), and defines it as follows [11]:

> "Human-Robot Interaction is a field of study dedicated to understanding, and evaluating robotic systems for use by or with humans."

A long-term goal of HRI is it to allow a more natural interaction between humans and robots, imitating human-human interaction [6]. Since the latter often requires communication, one has to think of ways on how to enable a natural dialog between humans and robots. Dependent on the respective field of application, we have to distinguish between different layers of communication. For instance, in an industrial context command based instructions are sufficient in order to control a machine remotely. For assistance systems that might not longer be the case. Thinking of robots as intelligent service providers, they should at least implement some kind of *Question answering* system. In more sophisticated applications robots should be able to interpret utterances, understanding what the intention was and consequently perform certain tasks. Furthermore this so called *Task-oriented dialogue* might be extended to a *Chat-oriented dialogue*, assuming that the robot is actively contributing to the dialogue rather than passively reacting to the user's input [12], [13].
The following section gives an introduction to Natural Language Processing, considering traditional and more advanced concepts. In the third section we will examine the possibilities a service robot can provide, having a special focus on the concept of *Robot-as-a-Service*. Finally we will see how both fields can be combined in order to obtain a more natural dialog between humans and robots.

II. Fundamentals of Natural Language Processing

Natural Language Processing describes the process of computationally understanding and manipulating human language [14]. Comparing machine language with human language, there are several differences. While human language has evolved over the years and is consequently rather unstructured, machine language is synthetic and structured. With human language one can express simple observations together with abstract concepts. Machine language by contrast is limited and often tailored to the respective purpose [15]. That is, a system is needed to translate one into the other.
Basically, such a system is composed of five components. The *Automatic Speech Recognition* (ASR) converts raw audio signals into a textual representation. This representation is then used as an input for the *Natural Language Understanding* (NLU) component, which interprets the utterance, for instance by matching it with a predefined set of rules, and finally converts it into a semantic representation. A *Dialog Management* (DM) component processes the semantic representation and uses some kind of knowledge base, that might be for example a task model, to put out a suitable response. Thereby the DM decides whether all preconditions for a certain task are satisfied or if there is any need for clarification. Based on that decision a respective textual response is generated by the *Natural Language Generation* (NLG) module. Finally, the *Text-to-Speech* component converts the textual output into an audio signal [16].
In the following we will focus on the NLU component since it represents one of the core components for creating a common ground between humans and robots or generally speaking embodied agents [17]. We will observe different methods for NLU, outlining advantages and disadvantages for their application in HRI.

A. Grammar-based NLU

In a grammar-based approach, the NLU component uses phrase spotting techniques that largely rely on syntax rules [18]. Due to performance reasons - one requirement to Natural Language Processing is that it should work in real time - these rules strictly follow the principles of a *context-free* grammar [19], meaning that there is only one nonterminal symbol S per rule, from which an arbitrary sequence of terminal and nonterminal symbols can be derived [20]. In a real-word scenario one could think of nonterminal symbols as abstract slots, that can be filled with concrete values [21]. Listing 1 shows a basic example, where a request for the weather can be applied to different cities and times. Note that rules are usually nondeterministic, so that a nonterminal symbol can be derived into multiple terminal values.

Listing 1. Basic example of a context-free grammar

```
$Loc → New York
$Loc → London
$Loc → Mainstreet in $Loc
$Time → today
$ROOT → How is the weather in $Loc $Time ?
```

Grammar-based approaches are applicable whenever user utterances can be easily transformed into syntactic representations [22]. Unfortunately, in HRI that is not always the case. Service robots often have to deal with elderly people or people with cognitive restraints, who might face problems when precisely formulating a request [6]. Since a grammar-based NLU is not capable of anticipating all linguistic constructions that might occur, communication can fail and HRI can be felt as unsatisfactory. To overcome the shortcomings of a grammar-based approach one could either use external modalities to support the dialogue (as discussed in section III) or implement more sophisticated approaches to NLU, such as using Machine Learning methods.

B. Deep Learning for NLU

Ambiguous utterances are a huge problem when it comes to Natural Language Understanding [19]. Rather than squeezing potentional utterances into a tightly tailored syntactic skeleton, one could think of ways to classify utterances and, based on that, approximating the users intention. Kim (2014) introduced Convolutional Neural Networks (CNN) to perform feature extraction on sentences in order to perform such a classification [23]. Being a specialisation of neural networks, CNNs consist out of multiple processing units (neurons), that are attached with a learnable weights. In consideration of a given activation function they produce an output by somehow computing a result given the weight and an input vector v [24]. This input vector is usually assembled from the calculation that previous neurons produce. Given the fact that there are multiple layers of neurons such a network is also called a *feedforward* neural network.

In CNNs there is no direct connection between the input layer and the output layer. Instead the output is calculated by using convolutions over the input layer. Thinking of the input as a matrix, one could imagine a convolution as a sliding window or a filter that is applied to the matrix. Figure 3 shows the concept for the given example of a $5x5$ matrix and a $3x3$ filter. Typically, a CNN comprises multiple convolution layers. During the training phase a CNN automatically learns the values of its filters. In Image Processing that could be for example the detection of edges in the first layer, then using these edges to recognise shapes in the second layer and so on [25].

Exactly like image pixels are represented by the input matrix in Image Processing, sentences serve as input in Natural Language Processing. Each row of the input matrix typifies a token in the sentence. That is usually a word but could also be a character. Consequently the filter is also a whole row having the same width as the input matrix [26]. With the aim to extract words, we can think of the filter as a vector of words, referring to the height of the vector (i.e. the amount of words) as the *region size*. For learning complementary features from the same regions, the CNN might use n filters of the same region size. As a result we obtain n vectors, containing the filtered words from the original sentence matrix. These vectors are also called *feature maps* as they represent the extracted features [27]. Based on the retrieved feature maps, another layer in the CNN performs a *1-max pooling* in order to record the n largest numbers from each feature map. In a simple example that could be the words with the highest occurrence rates. In a next step the *penultimate layer* concatenates the vectors gained from the 1-max-pooling and finally classifies the sentence using a softmax function [28].

The overall goal of using CNNs in text classification is to get a basic idea of what the text is about. As described in the preceding paragraph this is done by detecting the most "important" words of a sentence. However, this approach does only work to some extent. In fact it is only applicable in chat situations where the exact wording of an utterance does not matter. Indeed, the topic extraction allows building up a common ground between dialog partners in HRI. Other Machine learning techniques, namely *Sequence-to-sequence learning*, aim to understand a specific intention but allow the utterance to contain a certain amount of noise [29].

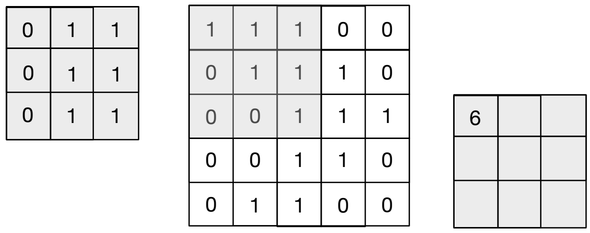


Figure 3. Functionality of a convolution. The matrix on the lefthand side is applied to the matrix in the center. The number of matches is than registered within another matrix.

III. Multimodality

When comparing Human-Robot Interaction (HRI) with Human-Computer Interaction (HCI), there are various similarities. Many algorithms that work for HCI are also applicable in HRI, for instance a speech recognition component is independent from the platform it is running on [30]. A major difference between both areas is that robots, and especially service robots, usually provide a great diversity on modalities, allowing a broader way of interaction. Next to several input channels, such as voice, gesture recognition, facial recognition and touch sensors, there are multiple output channels, some of which are motors, joints or LED displays [31]. Usually one would speak of *Degrees of Freedom* considering the ability of a robot to make movements, gestures or other expressions [32]. In traditional applications a degree of freedom corresponds with the joints of the robot. A robot arm for instance has five to six degrees of freedom, since it uses its joints (shoulder, elbow, wrist, fingers) to navigate trough a three-dimensional space and can additionally perform yaw, pitch and roll movements [33].

Furthermore, one could consider a robot's connection to the internet to be another input modality [34]. In order to perform tasks like perception, planning, control and particularly natural language understanding, robots need to have some kind of knowledge base. By querying large-scale knowledge systems, examples include the Google knowledge graph or IBM Watson, it is possible to conceptualise a dialogue manager that uses information from the semantic web. For this purpose contextualisation achieved by using information enrichment techniques [6]. To this extent we can embed a robotic system into a Service-Oriented Architecture (SOA), that comprises multiple services, such as knowledge bases or also cloud based cognition providers (cf. Microsoft Project Oxford [1]). Referring to the well-known concepts of *Software-as-a-Service* or *Hardware-as-a-Service* one could also speak of *Robot-as-a-Service* [35], meaning that the robot is designed and implemented to serve as a all-in-one SOA unit, requiring the robot operate as a client. That is, the robot needs to implement several interfaces to communicate with other, loosely coupled components and services.

[1] https://azure.microsoft.com/en-us/services/cognitive-services/

In order to combine both worlds – Natural Language Processing and Multimodality –, we have to leverage Natural Language Processing as an input modality, defining an architecture which processes inputs to actuate output modalities. In what follows we will introduce a basic concept for such an architecture.

IV. Embodied Cognitive Architecture for HRI

As an engaging social interaction is of great significance in most fields of application where service robots are situated in, a cognitive understanding is crucial for them to show appropriate, adaptive and contingent behaviours [36]. Using domain-general structures and processes as well as flat semantics, *Cognitive Architectures* can model the concept of cognition. Corresponding implementations are typically comprised of three layers. A knowledge base, usually specified as long-term or short-term memory, stores all the information that the agent has been given. That might be predefined concepts (beliefs, goals) or current information received from the input modules. Next, there is some semantic representation of the concepts that are stored in the knowledge base, allowing them to be gathered in a data structure. Based on this structure the agent can draw on the concepts received during interaction in order to perform a suitable reaction [37]. There are several implementations of cognitive architectures including ACT, Soar, ICARUS and Prodigy [38]. Since it mainly focuses on its usage in HRI, we will have a comprehensive look at ACT or more precisely the consecutive successor called ACT-R/E.

With the aim to produce human-like intelligence in an embodied context, ACT-R/E (Adaptive Character of Thought-Rational / Embodied) consists of five loosely coupled modules that represent a part of human recognition each [39]:

A. Declarative Module

The declarative module represents the knowledge base in the ACT-R/E architecture. Information is stored in chunks and managed by the module. Moreover additional meta data, such as the access frequency or situational contexts, helps the module to activate certain chunks, providing contextual information.

B. Intentional Module

The intentional module is responsible for controlling a task-oriented interaction. For instance, in task-oriented dialogue situations, the module ensures the appropriate response for the incoming request.

C. Imaginal Module

Particularly appreciable in ad-hoc interruptions of a dialogue, the imaginal module is capable of storing intermediate information, i.e. the system can revert to information that was gained in a former interaction.

D. Temporal Module

The temporal module serves the purpose to establish some sense of time within an interaction. That is, the system can distinguish between recent and past events and react correspondingly.

E. Procedural Module

The procedural module conjuncts the aforementioned modules by querying their current state and processes the information using its procedural knowledge. The procedural knowledge is mainly represented by production rules and can either be predefined or learned using reinforcement learning [40].

V. Conclusion

While there are myriads of approaches to HRI, Natural Language can be considered as one of the most wanted and most applied techniques, especially in the area of service robots. Yet one has to rethink the traditional methods of Natural Language Processing in order to adjust them to the needs of human dialog partners. We have seen that traditional approaches to Natural Language Understanding might confront a lack of coverage considering the broad diversity of slightly modified utterances a user can make. Machine Learning algorithms can provide a solution to this since they can recognise user intentions independent from their syntactic shape.

Moreover we can make use of the various modalities a robot supplies in order to support a dialog with additional information. For instance a robot can use its sensors to interact with a user, utilise its links and joints to visually explain things or even display images on a monitor or a projector. When connected to a cloud service, a robot can query semantic knowledge bases, enabling it to build a common ground with the user during conversation. That works particularly well in conjunction with things like topic extraction.

Combining both, the ability to understand natural language and the possibility to react sensibly (either verbal or nonverbal), we need to implement a cognitive architecture on the robot, allowing the robot to receive information in different input channels, process them and produce a corresponding output. We introduced ACT-R/E as one of the most advanced architectures in an embodied context.

Future challenges in combining Natural Language Processing with HRI will include further research in deep learning methods to improve semantic segmentation [41], a more precise representation of real world scenarios [42], as well as scaling the system to the complex structure of human language to cover more domains within a single conversation [43].

REFERENCES

[1] M. Rosheim, *Robot Evolution: The Development of Anthrobotics*, ser. Wiley interscience publication. Wiley, 1994.

[2] T. Ziemke, "Are robots embodied?" in *Lund University Cognitive Studies*, 2001, pp. 75–83.

[3] I. Asimov, *I, Robot*, ser. Robot series. Bantam Books, 1950.

[4] K. Dautenhahn and J. Saunders, *New Frontiers in Human-robot Interaction*, ser. Advances in interaction studies. John Benjamins Pub., 2011.

[5] C. Breazeal, "Social interactions in hri: The robot view," *Trans. Sys. Man Cyber Part C*, vol. 34, no. 2, pp. 181–186, May 2004.

[6] F. Alonso-Martín, A. Castro-González, F. J. F. de Gorostiza Luengo, and M. Á. Salichs, "Augmented robotics dialog system for enhancing human–robot interaction," *Sensors (Basel, Switzerland)*, vol. 15, no. 7, pp. 15 799–15 829, 07 2015.

[7] J. A. López and F. Cuéllar, "Robotman: Security robot for human-robot interaction inside malls," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 410–410.

[8] W. Barfield, *Cyber-Humans: Our Future with Machines*, 1st ed. New York, NY, USA: Copernicus, 2015.

[9] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robot. Automat. Mag.*, vol. 19, no. 2, pp. 98–100, 2012.

[10] M. Koschate, R. Potter, P. Bremner, and M. Levine, "Overcoming the uncanny valley: Displays of emotions reduce the uncanniness of humanlike robots," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 359–365.

[11] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: A survey," *Found. Trends Hum.-Comput. Interact.*, vol. 1, no. 3, pp. 203–275, Jan. 2007.

[12] R. E. Banchs, R. Jiang, S. Kim, A. Niswar, and K. H. Yeo, "AIDA: artificial intelligent dialogue agent," in *Proceedings of the SIGDIAL 2013 Conference, The 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 22-24 August 2013, SUPELEC, Metz, France*. The Association for Computer Linguistics, 2013, pp. 145–147.

[13] D. Spiliotopoulos, I. Androutsopoulos, and C. D. Spyropoulos, "Human-robot interaction based on spoken natural language dialogue," in *in Proc. European Workshop on Service and Humanoid Robots, 2001*, 2001, pp. 25–27.

[14] G. G. Chowdhury, "Natural language processing," *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 51–89, 2003.

[15] A. R. Wagoner and E. T. Matson, "A robust human-robot communication system using natural language for HARMS," in *The 10th International Conference on Future Networks and Communications (FNC 2015) / The 12th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2015) / Affiliated Workshops, August 17-20, 2015, Belfort, France*, 2015, pp. 119–126.

[16] U. Tiwary and T. Siddiqui, *Speech, Image, and Language Processing for Human Computer Interaction: Multi-modal Advancements*, ser. Premier reference source. Information Science Reference, 2012.

[17] A. Hartholt, T. Russ, D. Traum, E. Hovy, and S. Robinson, "A common ground for virtual humans: Using an ontology in a natural language oriented virtual human architecture," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, and D. Tapias, Eds. Marrakech, Morocco: European Language Resources Association (ELRA), 2008.

[18] H. Hofmann, U. Ehrlich, S. Reichel, and A. Berton, "Development of a conversational speech interface using linguistic grammars," in *5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Adjunct Proceedings*, 2013.

[19] G. van Noord, "Grammar-based natural language understanding," in *Priority Programme Language and Speech Technology*, 2003.

[20] T. Kretschmer, "Theoretische informatik," Vorlesungsskript, 2015.

[21] B. MacCartney. (2014, July) Understanding natural language understanding. [Online]. Available: https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf

[22] D. DeVault. (2013) Natural language understanding in dialog systems. [Online]. Available: http://projects.ict.usc.edu/nld/cs599s13/LectureNotes/cs599s13dialogue1-30-13.pdf

[23] Y. Kim, "Convolutional neural networks for sentence classification," *CoRR*, vol. abs/1408.5882, 2014.

[24] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[25] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *CoRR*, vol. abs/1511.08458, 2015.

[26] P. Xu and R. Sarikaya, "Contextual domain classification in spoken language understanding systems using recurrent neural network," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 136–140.

[27] Y. Zhang and B. C. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *CoRR*, vol. abs/1510.03820, 2015.

[28] F. Zhai, S. Potdar, B. Xiang, and B. Zhou, "Neural models for sequence chunking," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, S. P. Singh and S. Markovitch, Eds. AAAI Press, 2017, pp. 3365–3371.

[29] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *CoRR*, vol. abs/1409.3215, 2014.

[30] G. Rigoll, "Multimodal human-robot interaction from the perspective of a speech scientist," in *Speech and Computer - 17th International Conference, SPECOM 2015, Athens, Greece, September 20-24, 2015, Proceedings*, ser. Lecture Notes in Computer Science, A. Ronzhin, R. Potapova, and N. Fakotakis, Eds., vol. 9319. Springer, 2015, pp. 3–10.

[31] O. Novanda, M. Salem, J. Saunders, M. L. Walters, and K. Dautenhahn, "What communication modalities do users prefer in real time hri?" *CoRR*, vol. abs/1606.03992, 2016.

[32] A. Gupta and S. Arora, *Industrial Automation and Robotics*. University Science Press, 2009.

[33] K. Richards, *Design Engineer's Reference Guide: Mathematics, Mechanics, and Thermodynamics*. CRC Press, 2014.

[34] A. Saxena, A. Jain, O. Sener, A. Jami, D. K. Misra, and H. S. Koppula, "Robobrain: Large-scale knowledge engine for robots," *CoRR*, vol. abs/1412.0691, 2014.

[35] Y. Chen, Z. Du, and M. García-Acosta, "Robot as a service in cloud computing," in *Proceedings of the 2010 Fifth IEEE International Symposium on Service Oriented System Engineering*, ser. SOSE '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 151–158.

[36] P. Baxter and J. G. Trafton, "Cognitive architectures for human-robot interaction," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 504–505.

[37] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cogn. Syst. Res.*, vol. 10, no. 2, pp. 141–160, Jun. 2009.

[38] K. Gluck and R. Pew, *Modeling Human Behavior With Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. Taylor & Francis, 2006.

[39] J. G. Trafton, L. M. Hiatt, A. M. Harrison, F. Tamborello, S. S. Khemlani, and A. C. Schultz, "Act-r/e: An embodied cognitive architecture for human robot interaction," *Journal of Human-Robot Interaction*, vol. 2, pp. 30–55, 01/2013 2013.

[40] J. Anderson R. and W.-T. Fu, "From recurrent choice to skill learning: A reinforcement-learning model," *Journal of Experimental Psychology: General*, vol. 135, no. 2, pp. 184–206, 2006.

[41] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodríguez, "A review on deep learning techniques applied to semantic segmentation," *CoRR*, vol. abs/1704.06857, 2017.

[42] H. Khayrallah, S. Trott, and J. Feldman, "Natural language for human robot interaction," in *Proceedings of the Workshop on Human-Robot Teaming at the ACM/IEEE conference on Human-Robot Interaction (HRI)*, 2015.

[43] A. Aghaebrahimian and F. Jurcícek, "Machine learning for semantic parsing in review," 2015.