

Survey on Spoken Dialogue Systems: User Expectations Regarding Style and Usability

Markus Berg, University of Wismar, Germany

Abstract

Most people already came into contact with Spoken Dialogue Systems, mostly in form of automated call centres. Our aim is the improvement of these systems. In this paper, we focus on the users' preferred style of interaction. We conduct a survey and evaluate the results, which will be the basis for our future work in the context of modelling natural language dialogue systems.

1. Introduction

In this paper we evaluate the results of a survey about the user expectations regarding the style of a dialogue system. We first give a short introduction into the area of dialogue systems and consequently place this work into the right context.

1.1. Dialogue Systems

„A spoken dialogue system enables a human user to access information and services that are available on a computer or over the Internet using spoken language as the medium of interaction" [1]. Apart from spoken language, also text-based input, as in chatbots, is possible. A spoken dialogue system (SDS) consists of five components: The speech recogniser transforms spoken language into text. It is then interpreted by a language understanding component. The dialogue manager interacts with the back-end and decides how to react on the user's request. Depending on this request and the result from the back-end, the system generates a question or an answer. The specific formulation is then created by the language generator and is eventually synthesised.

The first text-based dialogue system was Weizenbaum's *ELIZA* [2] – a simulated psychologist. Later, in Winograd's *SHRDLU* [3], you could move blocks in a simulated world by natural language commands. Today, most dialogue systems are used in the context of automated call centres and speech based information systems. These systems offer telephone-based, automated services like pre-qualification (determine the correct department for the customer's problem), bus timetable information, or topping up your pay-as-you-go phone. While

some of these systems still only use a key-based navigation (DTMF), others also support speech recognition. Because only few systems are able to understand full sentences, many produce long and tedious dialogues. At the same time, most systems always use the same formulations and a telegraphic and non-human-like style.

1.2. Goal and Context of the Survey

Our aim is the improvement of spoken dialogue systems. So we need to know what users think of current dialogue systems and what they wish for the future. Despite some authors claim, that we don't need natural and human-like dialogue systems, we think this will be the future. It may be the case that today, human-like systems create many difficulties, because the analysis of natural language utterances is a complex task. It is easier to process restricted commands like in simple command-and-control systems. Users also adapt to this limited interaction style and are able to use it efficiently, but we don't believe that they wouldn't be happy with more human-like systems, if they existed.

We want to find out, which type of dialogue system users like most. Our assumption is that, in general, users prefer systems with human-like capabilities and a natural style. We also imagine that there are no clear relations between preferred style and age or background. We believe that every user is different and that we need systems that adapt to the users with regard to style and formulation.

2. Previous and Related Work

Berg et al. [5] conducted a Wizard of Oz experiment to find out whether and how people want to talk to computers in terms of solving everyday tasks like writing an email with only having a speech interface at hand. It has been shown that most people – independent from age, gender, experience or profession – stick to the Graphical User Interface (GUI), i.e. they use commands that they remember from the GUI. Moreover, we have learned that most of the participants categorise natural language (full sentences) as helpful. In case of unease or in the

need to solve a complex task, users switched from short commands to natural language. Furthermore, when explicitly offering the possibility to naturally interact with the computer, the participants were surprised and stated that this would be much easier than the command-oriented control.

Kruijff-Korbayova and Kukina [6] compared “*system evaluation judgments and input style alignment of users interacting with an in-car dialogue system generating output in personal vs. impersonal style.*” They did not find differences regarding the users’ attitudes to the system among the styles.

Brennan and Oheari [4] figured out whether systems with an anthropomorphic message style appear as being more intelligent than systems with a telegraphic style. They did not find any evidence for this theory. But as users align to the system’s style and because the processing of anthropomorphic utterances is much more difficult, they concluded that this style may not be preferable.

Brown and Levinson [7] claim that trying to fulfil the human need for self-esteem and respect from others is called politeness. This need can also be fulfilled by computers. So, Hjalmarsson and Edlund [8] analysed the effects of human-likeness in utterance generation. As a part of their study, they asked users to rate the importance of human-likeness, politeness, efficiency, intelligence and understanding.

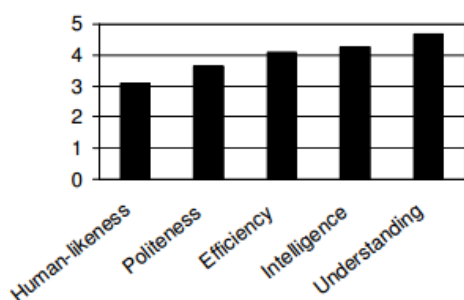


Figure 1: Mean importance ratings [8]

Although human-likeness and politeness have the lowest score, they still yield to averages above 3 of 5 points. This underlines the importance of these features. Of course, understanding is the most important feature, because a friendly, but badly working dialogue system is useless.

Dautenhahn et al. [9] presented a study about people’s attitudes towards robot companions for the home. They found out that “71% of subjects would want a robot companion to communicate in a very human-like or human-like manner”. Edlund et al. [10] noted that “some users happily respond to greetings and social system utterances, whereas others never respond to them at all.”

3. Questions and Method

The survey is completely web-based and consists of 18 questions that can be seen in figure 2. It is structured into four main categories:

- General questions (3)
- Experience with SDS (5)
- Opinions regarding style (6)
- Examples regarding style (4)

In the first category, we ask general questions regarding sex, age and computer skills. This is followed by questions about what users dislike about current spoken dialogue systems and how satisfied they are.

1. Gender
2. Age
3. Computer Skills (4 options)
4. Have you ever used a Speech Dialog System (talking computers, e.g. hotlines)?
5. How satisfied have you been with the operation/usability? (0..5)
6. What did you dislike? (7 options)
7. What needs to be improved in today's systems or how are systems from television (e.g., Star Trek) better? (5 options)
8. Do you think that speech understanding systems are (in future) beneficial or do you think that these systems are superfluous? (4 options)
9. Independent of the current technology: Would you like to instruct the computer in short commands or would you rather use full sentences? (3 options)
10. Should a computer speak in full sentences or do you prefer short keywords as answers? (2 options)
11. What do you think of open questions? (questions without predefined answer like "How may I help you?") (2 options)
12. What do you think of directed dialogs? ("Say balance, change plan, or failure report") (3 options)
13. Should a computer be polite? Should it greet you and say "please" and "thanks" (3 options)
14. Are you polite yourself when talking to a computer? (Do you say please and thanks?) (3 options)
15. If you want to tell the computer to switch the light on, what do you say? (4 options)
16. If you tell a co-worker to switch the light on, what do you say? (4 options)
17. Which dialogue do you think is more pleasant? (2 options)
18. You want to know how the weather will be. Which dialog system would you use? (3 options)

Figure 2: Question overview

The third category focuses on what users expect from future systems, if they are interested in polite man-machine-dialogues and if they prefer human-like sentences to short machine-like statements. In the last category we present some exemplary dialogues in different styles and let the users decide which one they like most.

4. Results

The survey has been completely answered by 194 participants. While the survey was active from 14 February until 23 July 2012, 86% of all answers have been given in February and 16% in March. It was advertised among friends, colleagues, students and LinkedIn groups. 85.6% of the survey has been taken in German and the remaining 14.4% in English. The mean time for completing the survey was 5m52s, resulting in 16 seconds per question.

In the following sections we will not give a complete overview about every detail, but instead highlight interesting results.

4.1. Socio-Demographic

The gender distribution is almost equal with a slight majority of male participants (55.2%). The majority is aged between 19 and 35 (19-25: 25.3%, 26-35: 47.4%). There were only 0.5% under 19 and 3.1% older than 65. 14.9% were between 36 and 49 and 8.8% between 50 and 65. The distribution of male and female participants is very homogenous throughout the age groups with a derivation of less than 5%. 41.2% of all participants are IT specialists or computer scientists, 50% use the computer at work on a daily basis and 8.8% use a computer only at home.

4.2. Experience & Satisfaction

90.7% of all participants claim to have used a SDS at least once in their life.

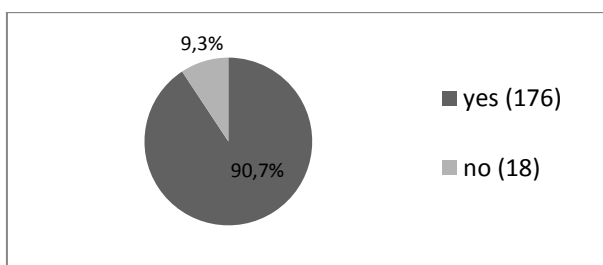


Figure 3: Have you ever used a SDS?

On a scale from 0 to 5, the mean satisfaction was 2.63 with a standard deviation of 0.86. Asking for the main problems with dialogue systems resulted in the following top three answers:

- The 'question-answer-game' was very tedious (62.8%)
- The system did not understand me (55.2%)
- The system offered too many choices (51.7%)

In relation to the last point, 45.3% claimed to have difficulties in identifying the correct menu item. Surprisingly, only 22.7% complained about systems that are only able to recognise numbers, while at the same time 38.4% complained about systems that only recognise single words instead of full sentences. After asking for the problems users have with current dialogue systems, we asked what should be improved and where fictional systems from TV are better. 60.8% voted for a better speech recognition. This result has been expected, because 55.2% stated, that the system often does not understand the user. It has to be mentioned, that the participants might not have been aware of the difference between speech recognition and natural language understanding. Although only 38.4% complained about systems, that only understand words, 57.4% voted for capabilities to understand full sentences. 79.5% even want to actively tell the computer their concern instead of answering lots of questions. Of course this ability includes the understanding of full

sentences. This can be interpreted as follows: While approximately 40% regard the lack of understanding full sentences as severe, 60% think it would be good if systems could understand full sentences and 80% wish for a technology that enables the users to actively tell their story. 29.5% voted for a better synthesised voice and only 1.7% thought that there is no need at all to improve current systems.

Afterwards, we asked if speech dialogue systems are beneficial. 26.8% said yes and 44.8% think they will be beneficial in the future, given that they work better. Alarmingly, every fifth participant claims that he does not need these systems today, nor will he in the future.

4.3. Preferred Style (Self Assessment)

In the third part of our survey, we asked users which style they prefer. 51.0% want to use full sentences ("How will the weather be in Paris tomorrow?"), while 34.5% prefer medium-length commands ("Weather in Paris for tomorrow"). The rest voted for short commands like "Weather, Paris, tomorrow".

But also the system can produce sentences. Here, 63.9% want the system to formulate full sentences ("I have found five flights. The next one, with flight number 376, leaves at 4pm, ...") instead of short enumerations and keywords ("Five results: Flight 376, 4pm; flight 874, 5pm").

57.7% prefer open-ended questions like "How can I help you?" to specific questions, that result in long menus ("Do you want A, B, C or D?").

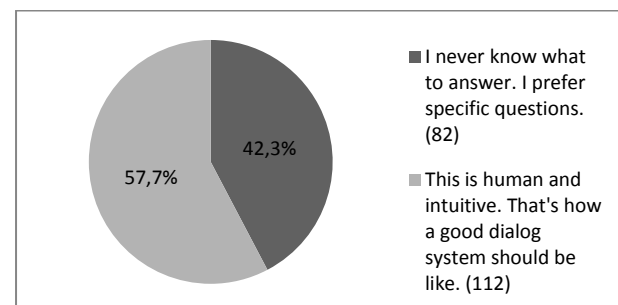


Figure 4: What do you think of open questions? (Questions without predefined answer like "How may I help you?")

About half of all participants want the system to be polite and want it to say please and thanks. 36% don't care and only 16% think this is unnecessary. Again, slightly more than 50% are polite when addressing the system, while 36% sometimes say please and 12% don't do it at all.

4.4. Preferred Style (Examples)

When addressing a computer with a command, users only use few words and seldom make use of polite formulations. This type of interaction is comparable with pressing a button to switch on the light. On the other hand, when addressing a co-worker, people mostly use very polite formulations. Edlund et al. [10] explain this behaviour with the interface

metaphor: *“Within the interface metaphor, the spoken dialogue system is perceived as a machine interface – often a computer interface. [...] Within the human metaphor, on the other hand, the computer is perceived as an interlocutor: a being with human-like conversational abilities”.*

Apart from instructions, we also wanted the participants to choose their favourite dialogue style. 71.1% preferred dialogue A in question 17 (see figure 6). It is very natural and human-like, starts with an open-ended question and allows the user to control the dialogue flow. But still, every fourth user likes the restricted dialogue style more, perhaps because they are used to it.

In the last question (see figure 7) the participants had to choose between three dialogue styles for getting weather information. Dialogue B was the most human-like, dialogue A was still more natural than most of today’s systems and dialogue C was the machine-like style that we know from current dialogue systems. 52.6% chose dialogue B as their favourite, while 38.1% voted for dialogue A. Only less than 10% prefer the system-initiative dialogue.

5. Analysis

Having this data, we can now infer relations between survey results and characteristics like age or gender.

5.1. Computer usage

We first want to know if the type of computer usage affects the wish for politeness of spoken dialogue systems. We also want to make sure that the big amount of computer scientists in our study does not tamper the results. Nevertheless, we still have to be aware of the fact that only technical oriented people participate in an online survey.

Politeness: We found out that 41.3% of the IT professionals say it is important, that the computer is polite (38.8% don’t care) as opposed to 54.7% of the non-IT people (30.9% don’t care), and 52.9% of the home users (47.1% don’t care). Many IT professionals are used to command lines and seem to see the computer as a tool. Regular users may be happier with a more personal computer that acts as a partner to help them fulfilling the task.

On the other hand, 9% of the IT professionals and 15% of the non-professionals state that they are always polite when speaking with a computer. Interestingly, only 6% of the hobby users address the computer in a polite way. At the same time, this group has the highest value in people saying, that they are sometimes polite (41%).

Open-ended questions: Open-ended questions are popular among IT professionals (63%) and hobby-users (65%). Only 53% of the non-IT people, who use the computer at work like this type of questions.

Full sentences: 60% of the computer scientists want to use full sentences, whereas only 44% percent of the people who use computers at work and 47%

of the hobby users want to formulate their requests in full sentences.

5.2. Gender

Politeness: For females, politeness is more important (54%) than for males (45%).

Open-ended questions: We could identify a male preference for open-ended questions (62% vs. 53%).

Full sentences: For males it is more important (67%) that the computer uses full sentences, than for females (60%). About half (51%) of both women and men also want to be able to use full sentences. 36% of the women and 34% of the men prefer medium-length sentences and slightly more men (15%) than women (14%) think words are enough.

5.3. Age

Because we only have 1 and 6 participants in the youngest and oldest category, we omit these in the following results and only regard ages from 19 through 65.

Politeness: Our hypothesis is that the older people are, the more do they appreciate politeness. This has not been verified, as the following results show:

- 19-25: 42,9%
- 26-35: 52,2%
- 36-49: 55,2%
- 50-65: 52,9%

Open-ended questions: We also thought that older people prefer open-ended questions, because this is more human-like. Again, there is no strong evidence for that. In fact, the favouring of open-ended questions seems to decrease with increasing age (until the age of 49). However, the age group from 50-65 has, as expected, the highest percentage.

- 19-25: 59,2%
- 26-35: 57,6%
- 36-49: 51,7%
- 50-65: 64,7%

Full sentences: Another hypothesis is that older people favour full sentences. We can see that the age group 50-65 indeed has a 7% higher preference for full sentences, while at the same time the 26-35 group likes full sentences 11% less than participants of the groups 19-25 and 36-49.

- 19-25: 69,4%
- 26-35: 57,6%
- 36-49: 69,0%
- 50-65: 76,5%

5.4. Analysis of the Examples

We now try to verify the answers by giving the participants examples, where they should choose the dialogue they like most.

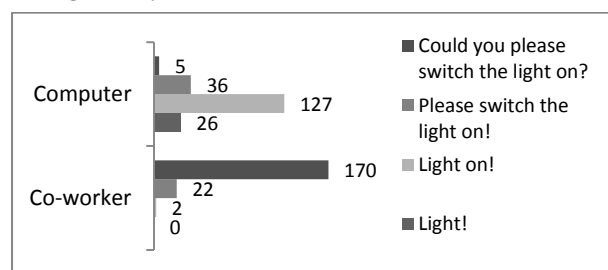


Figure 5: If you want to switch the light on, what do you say?

We already learned, that when it comes to actions that control devices, like switching lamps (see figure 5, question 15), people mostly use short commands. Apart from the mere frequencies, we also analysed contingency tables. Only 35% of the participants who claimed always to be polite said *please* and 4% said *could you please*. Similarly, 69% of the people preferring full sentences only say *light* or *light on*. In question 17, that you can see in figure 6, we presented the participants two dialogues. One being very human-like with the ability to understand full sentences, and another more machine-like and command-based.

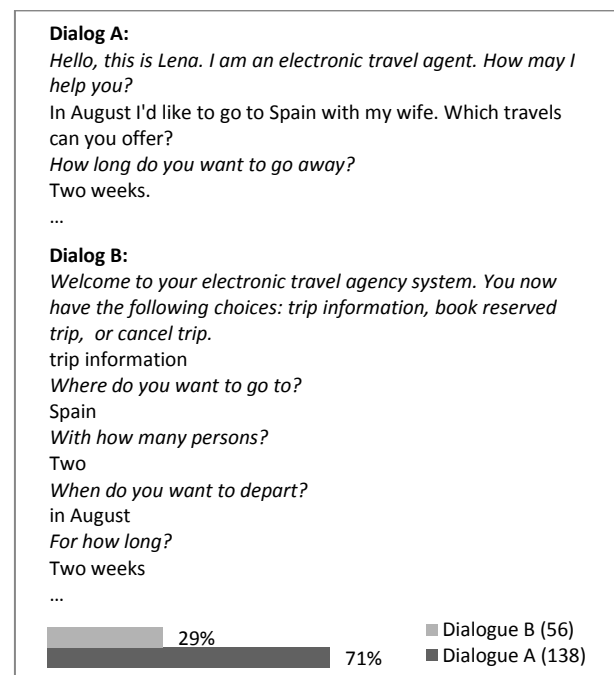


Figure 6: Which dialogue do you think is more pleasant?

We analysed the results in relation to the following features.

a) Full sentences:

- 83% of the participants, who prefer full sentences, like dialogue A most.
- Surprisingly, only half of the participants, who prefer short commands, like dialogue B most.

b) Many tedious questions:

- 72% of the participants, who don't like tedious question-answer-dialogues, prefer dialogue A. Also 70% of the people who actively want to tell the computer their concern instead of answering lots of questions prefer dialogue A.
- Interestingly, also 70% of the participants who are not annoyed by the tedious question-answer dialogues, like dialogue A most.

c) Open-ended questions:

- 81% of the participants, who like open-ended questions, prefer dialogue A.

- Surprisingly, 57% of the participants, who claimed not to like open-ended questions, still prefer dialogue A. Only 19% of the people, who like open-ended questions, prefer dialogue B.

d) Directed dialogue:

- 75% of the participants who have problems with directed dialogues ("*Say balance, change plan, or failure report*") or just think they are tedious prefer dialogue A.
- On the other hand, only 53% of the people, who like this dialogue-style, prefer dialogue B.

Now we ask the participants to choose the best out of three dialogues (see figure 7). This time we also offer a version with medium-length commands.

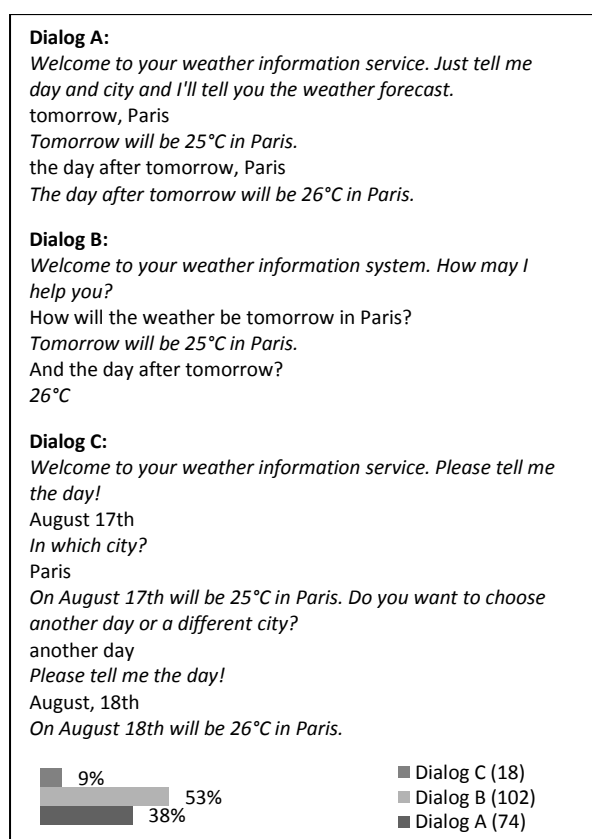


Figure 7: You want to know how the weather will be. Which dialog system would you use?

Furthermore, we can check if the answers are consistent, i.e. if people who prefer the more natural dialogue from the last question, necessarily prefer the natural dialogue in this question. Again, we checked with reference to different relations:

a) Full sentences:

- 27% of the participants, who prefer full sentences, like dialogue A most. 66% of them prefer dialogue B.

b) Many tedious questions:

- 53% of the people, who don't like systems that ask lots of questions, prefer dialogue B and 41% prefer dialogue A.

c) Open-ended questions:

- 60% of the people who prefer open-ended questions, like dialogue B and 32% prefer dialogue A.
- Interestingly, almost half of the participants (43%), who dislike open-ended questions prefer dialogue B.

5.5. Familiarity with Dialogue Systems

Finally, we put a special emphasis on the opinion of new users, i.e. participants who have never used a dialogue system before. We made the following observations:

- 77% prefer open-ended questions
- 72% prefer dialogue A in question 17 and 78% prefer dialogue A or B in question 18
- 72% are between 19 and 35, 67% are female, and 61% use a computer daily
- 61% say commands like *light on* when it comes to control situations
- Only 28% prefer directed dialogues with many question-answer-pairs

6. Conclusion

In this survey we have identified problems of current SDS and analysed how users would like to interact with them. 91% of all participants have already used a SDS. Interestingly, 72% stated that SDS are already beneficial today or believe that they will be beneficial in the future if they are improved. This underlines the importance of this area of research both for academia and industry. But there are still some drawbacks that need to be improved. Almost 63% do not like the tedious way of having to answer lots of questions, before being able to get the relevant information. Instead, almost 80% want to actively describe their problem.

This is at odds with only 51%, who say that they would like to use full sentences and 64% who want the system to make use of full sentences. 58% prefer open-ended questions. We assume that the participants are not aware, that these features are needed in order to realise the above mentioned characteristic.

Consequently, after asking self-assessment questions, we presented the participants example dialogues, to verify their claims. We found out, that even more people favour human-like systems, than indicated in the questions before. 71% like the most natural dialogue in question 17, which makes use of an open-ended question and full sentences. In question 18 only 9% favour the directed dialogue. Also, half of the people having said, that they only like short commands, surprisingly like the very natural dialogue most. Also 70% of the people, who claim not to be annoyed by long question-answer pairs, still prefer the natural dialogue. Even 57% of the participants,

who claim not to like open-ended questions, prefer exactly this type of dialogue. Moreover, only 53% of the people that prefer directed dialogues really choose the directed dialogue as the best one. When we take a look at users who have never used a SDS before, we can as well clearly identify a preference for open questions (77% as opposed to 58%).

Only 16% regard polite systems as unnecessary. The rest either favours politeness or doesn't care. This shows that polite systems are important for most people, although only 50% are polite themselves when addressing the system. This also confirms Edlund [10], who says that some people respond to greetings and social utterances, while others don't.

Although anthropomorphic systems may not seem more intelligent, we can clearly see a preference for those systems, i.e. systems with open-ended questions, understanding and generation of full sentences, and the use of polite formulations. This is in line with the results of Dautenhahn et al. [9], who found out that 71% wish for a human-like communication with robots. However, we could not identify a clear relation between the wished style and age, gender, or IT skills.

Bibliography

- [1] K. Jokinen and M. F. McTear, *Spoken Dialogue Systems*. Morgan & Claypool Publishers, 2009.
- [2] J. Weizenbaum, "ELIZA - A computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, Jan. 1966.
- [3] T. Winograd, *Understanding Natural Language*. New York: Academic Press, 1972.
- [4] S. E. Brennan and J. O. Ohaeri, "Effects of message style on users' attributions toward agents," in *Conference companion on Human factors in computing systems*, 1994.
- [5] M. Berg, P. Gröber, and M. Weicht, "User Study: Talking to Computers," in *3rd Workshop on inclusive eLearning*, 2010.
- [6] I. Kruijff-Korbayová, C. G. O. Kukina, and J. Schehl, "Generation of output style variation in the SAMMIE dialogue system," in *Proceedings of the Fifth International Natural Language Generation Conference*, 2008.
- [7] P. Brown, S. C. Levinson, and J. J. Gumperz, *Politeness: Some Universals in Language Usage*. Cambridge University Press, 1987.
- [8] A. Hjalmarsson and J. Edlund, "Human-Likeness in Utterance Generation: Effects of Variability," in *Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems*, 2008.
- [9] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry, "What is a robot companion - friend, assistant or butler", 2005.
- [10] J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, "Towards human-like spoken dialogue systems," *Speech Commun.*, vol. 50, 2008.

Author



Markus Berg, M.Eng.
University of Wismar,
Faculty of Engineering,
Deptmnt. of EE and CS
markus.berg@hs-wismar.de
<http://mberg.net>