Natural Language Input for In-Car Spoken Dialog Systems: How Natural is Natural?

Patricia Braunger, Wolfgang Maier, Jan Wessling, Steffen Werner
Daimler AG
Sindelfingen
Germany
{patricia.braunger, wolfgang.mw.maier}@daimler.com
{jan.wessling, steffen.s.werner}@daimler.com

Abstract

Recent spoken dialog systems are moving away from *command and control* towards a more intuitive and natural style of interaction. In order to choose an appropriate system design which allows the system to deal with naturally spoken user input, a definition of what exactly constitutes *naturalness* in user input is important. In this paper, we examine how different user groups naturally speak to an automotive spoken dialog system (SDS). We conduct a user study in which we collect freely spoken user utterances for a wide range of use cases in German. By means of a comparative study of the utterances from the study with interpersonal utterances, we provide criteria what constitutes *naturalness* in the user input of an state-of-the-art automotive SDS.

1 Introduction

In the automotive area, speech interfaces have continously gained importance in recent years. Current spoken dialog systems (SDS) are expected not to be restricted to a *command-and-control*-style interaction, in which functions are invoked by the user by speaking fixed key phrases. Instead, they are expected to accept *natural* input from the user, i.e., to understand the user without imposing restrictions on how he has to formulate queries.[1]

A definition of what exactly constitutes *naturalness* in user input is important, not only in order to precisely understand user expectations, but also, and especially, in order to choose an appropriate system design which allows the system to deal with flexible user input and spontaneous speech phenomena (as described by Skantze (2007)), and to facilitate the design of meaningful system evaluation.

Since interpersonal interaction is the most natural form of interaction, it is often taken as a baseline for the development of an intuitive and natural human-machine interaction (Bonin et al., 2015). However, earlier work shows that human speech is strongly influenced by the assumptions that a speaker has about his interlocutor, e.g. (Branigan and Pearson, 2006), and also by individual properties such as age, e.g. (Möller et al., 2008; Bell, 2003). In conclusion, naturalness in user input cannot simply be equated with interpersonal speech and different user groups may have a different understanding of what is natural and intuitive. To the best of our knowledge, there are no studies investigating what exactly constitutes naturalness in user input.

In this paper, our aim is to answer the question of which kind of utterances the natural language understanding component of an SDS must be able to understand from a user perspective. Thereby, characterizing the capabilities of a dialog management, as done by Bohlin et al. (1999) (cf. TRINDI tick-list), is not enough – a thorough characterization of the characteristics of natural language user input is needed. In order to achieve this, we conduct a study in which we collect free user utterances for an in-car SDS in German. By means of a comparative analysis with interpersonal utterances, we first show to which extent utterances used for system interaction share properties with interpersonal utterances. Second, we examine to which extent different user groups speak differently in terms of naturalness.

The remainder of the paper is structured as follows. In section 2, we review previous literature which has aimed at defining naturalness of user input and describing natural language utterances respectively. The following section 3 we introduce

[1] Also, it is expected, that systems answer *naturally* to the user. However, a discussion of system output is beyond the scope of this paper.

Proceedings of the SIGDIAL 2017 Conference, pages 137–146,
Saarbrücken, Germany, 15-17 August 2017. ©2017 Association for Computational Linguistics

our study design. Section 4 presents the evaluation of the study, in section 5 the results are discussed and section 6 concludes the article.

2 Towards a Definition of Naturalness

In general, natural language is human language and therefore different from artificial languages which are especially created for specific purposes, e.g., computer languages. In this sense, spoken dialog systems always make use of natural language. This also applies to command-and-control systems. However, the term *natural* is often used as a qualifier of the abilities of the natural language understanding (NLU) and natural language generation (NLG) modules of an SDS.

A general definition of naturalness in this sense is given by Berg (2013), who calls SDS *natural* if their language behavior is as human-like as possible. Many authors refer to this definition of *natural language* when they demand a more natural human-machine interaction, see, e.g., (Edlund et al., 2008).

The literature that investigates the naturalness of spoken user input, which is the focus of our work, can be split into three groups.

Literature in the first group describes the users' speaking style by means of labels like *natural* and *command*. Hofmann et al. (2012), e.g., conduct a web-based study to find out how users would interact with internet services using speech. They classify the observed speaking styles into *natural*, *command* and *keyword style*. They state that *natural* reflects the way humans communicate among each other and that the *command* and *keyword style* is related to state-of-the-art human-machine interaction. Berg et al. (2010) use similar labels with a different meaning. They classify utterances collected from a human-machine interaction study into *commands*, *phrased commands* and *natural language*, whereas *commands* is used similar to *keyword style* of Hofmann et al. (2012) and *natural language* utterances consist of full sentences including phrases of civility and filler words. Similarly, in the study of Berg (2012), speaking styles are classified into *full sentences*, *medium-length commands* and *short commands*. White et al. (2014) and Pang et al. (2011) investigate written web search queries. They classify information seeking queries into *keyword queries* and *natural language questions*. *Natural language questions* are defined as utterances beginning with a question indicator, such as *what* and *do*, and ending with a question mark.

The second group consists of literature which (linguistically) analyzes spoken user input style. Braunger et al. (2016), e.g., compare crowd-sourced natural language user input in terms of sentence constructions. They conclude that if people speak freely to an SDS, they mostly use an imperative style. Winter et al. (2010) collect naturally spoken utterances and quantify their complexity and variety. They use context information as a qualitative measurement for classification, classifying the utterance content into three categories: *information data*, *context relevant words* and *non-context relevant words*. They find that users tend to repeat similar utterance patterns composed from a limited set of different words.

Thirdly, we find work which concenctrates on the differences between human-human and human-machine communication. Guy (2016) shows that voice queries are closer to natural language than written queries. He builds two natural language models, one based on a corpus representing classic formal language and one based on a corpus representing a more colloquial web language. For measuring the similarity to a natural language model he used perplexity. He concludes that voice queries are still far from natural language questions. The authors of (Hayakawa et al., 2016) compared direct human-human dialogs to dialogs that are mediated by a speech-to-speech machine translation system. They found that in machine mediated conversation speakers use less words than in direct human-to-human communication. In (Pang and Kumar, 2011) written natural language questions posed as web search queries are compared to a natural language sample of questions posted by web users on a community-based question-answering site. Since written text tends to be structurally complete Pang et al. (2011) measure naturalness by means of the probability mass of function words.

A more intuitive and natural interaction with SDSs presupposes understanding naturally spoken user utterances. In order to choose an appropriate system design which allows the system to deal with naturally spoken utterances, a definition of what exactly constitutes naturalness in user input is necessary. Recent research in this area only focuses on the question whether users speak naturally or in a command-/keyword-based way to a speech system, whereby *naturalness* is equated with human-directed speech, e.g. (Hofmann et al., 2012; Pang and Kumar, 2011; Berg, 2012). The criteria mentioned for natural, human-directed speech are full sentences, civility, filler words and a higher number of words. Since natural is what people intuitively use, *natural language input* cannot simply be equated with interpersonal speaking style. Even though different studies found that a speaker's language behavior is influenced by beliefs about an interlocutor, cf. (Branigan and Pearson, 2006; Branigan et al., 2010; Bell, 2003) and researchers have many intuitions about the differences between human-machine and human-human communication, interpersonal speaking style is often taken as a baseline for naturalness as can be seen from the discussed literature and it has not been examined to which extent the criteria mentioned for naturalness characterize naturally spoken utterances towards state-of-the-art SDS. There exist only a few empirical studies which investigate the differences. These research works focus either on dialog issues such as turn-taking, e.g. (Doran et al., 2001), or on lexical alignment, e.g. (Branigan et al., 2011), but not on natural language input towards SDS in a car environment.

The way people address the system is not only influenced by their beliefs about the system but also by individual properties such as age or gender. Work in this area of research has been done by Bell (2003) who found that individual differences in speaker behavior are significant and by Möller (2008) who found that younger users differ from older users in the way they speak with a smart-home system. The observations show that different user groups may have a different understanding of what is natural and intuitive. Therefore, user profiles must be considered when defining natural language input.

3 Study Design

To the best of our knowledge, there are no data answering the question to which extent naturally spoken user input towards SDS differ from human-directed speech and what exactly constitutes naturalness in user input. We therefore conduct a study to examine how different user groups would naturally speak to an actual in-car SDS and how they would speak to their passenger.

In the following, we explain the experimental setup and procedure of the study.

3.1 Participants

The study is targeted at younger and elder German adults with different SDS experience and a valid driver's license. In total, 45 subjects participated in the study. 46% of them were female and 54% were male. The average age was 39.5 years (standard deviation SD: 13.5). 55.6% of the participants were aged between 20-39 years, 26.6% were 40-59 years old and 17.8% were older than 60 years. 27% were experienced in the use of spoken dialog systems; 74% had little to no experience with speech-controlled devices.

3.2 Experimental Design

The study was split into two sessions and each participant encountered both conditions (*within-subject design*). In the one session the participants had to talk to their front passenger who performed the requested action. In the other session the participants were asked to interact with an in-car spoken dialog system. According to Möller (2008; 2005) we decided to conduct a Wizard of Oz (WOZ) experiment. This method is less time consuming and less costly. In a WOZ experiment a human operator (wizard) simulates the behavior of an intelligent computer application whereby the human believes to be interacting with a fully automated prototype (Dahlbaeck et al., 1993). Within each session the participants were asked to solve twelve tasks typically performed in a car:

1. Listen to radio station SWR3
2. Play Michael Jackson Greatest Hits
3. Navigate to Stieglitzweg 23 in Berlin
4. Call Barack Obama on his mobile phone
5. Set temperature to 23 degrees
6. Send a text message to brother
7. Weather in Berlin today
8. Date of the European Football championship final game
9. Population of Berlin
10. Score VfB Stuttgart against FC Bayern
11. Cinema program in Berlin today
12. Next Shell gas station

The tasks consist of six non-information seeking tasks (1-6) and six information seeking tasks (7-12).

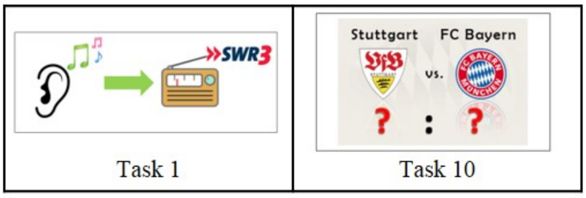

Figure 1: Task description

3.2.1 System Simulation

The system behavior was simulated with the help of the SUEDE tool (Klemmer et al., 2000). The system behavior was designed such as in an actual Mercedes-Benz E-class. The system directly provided the information requested or activated the appropriate function whereby the user input resulted in a visual and acoustic system feedback. With user input for Task 1), for example, the radio program started playing and the screen provided information on the current radio station.

3.2.2 Task Description

The tasks were presented by pictures in paper form. Different studies, e.g., (Bernsen et al., 1998; Tateishi et al., 2005), report from priming effects when using text-based task descriptions. As pictures do not bias the subjects by putting words into their mouths, the participants were shown pictures that describe the tasks. The tasks were pre-tested with friendly users to find out if the desired situation was put in the user's mind. Examples for the task descriptions are given in Fig. 1.

3.2.3 Driving Simulation Setup

Since we want to find out how users naturally interact with a spoken dialog system while driving, we put the participants in a simulated driving situation. The participants were sitting on the driver's seat in a car which was placed in front of a canvas onto which the driving simulation was projected, such as done by Hofmann et al. (2014). They were shown a driving simulation where they were driving behind a car. Their task was to brake if and only if the preceding vehicle brakes. The driving simulation setup is illustrated in Fig. 2.

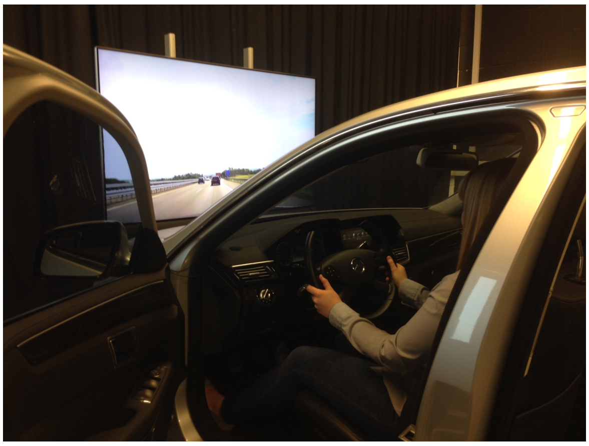

Figure 2: Driving simulation setup

3.3 Procedure

The overall procedure of the experiment was as follows. First, the participants were informed about the procedure. The participants were told that they have to orally solve tasks while driving and they were shown the graphically depicted tasks. The participants had to verbally interpret the tasks. In order to prevent wrong interpretations we gave assistance, where necessary. As for the session with the passenger, they were told that the passenger provided the information requested or activated the appropriate function. As for the system session, they were told to speak freely to the system. They had to activate the speech recognition via speaking the phrase "Hallo Auto" (eng. "Hello Car"). Afterwards, the participants got to know the driving simulation in a test drive lasting about three minutes. The instructor was sitting on the passenger seat. The instructor showed the task presentation pictures randomly while the participant was driving. The tasks were permuted to avoid order effects.

4 Evaluation and Results

In total, we collected 1.080 utterances; 540 system-directed utterances and 540 human-directed utterances. The utterances were manually transcribed and automatically analyzed. The transcription exactly matched the spoken utterances. The analysis included Part-of-Speech (POS) Tagging and Parsing with *SpaCy*.[2] The part-of-speech-tagger uses the Google Universal POS tag set of Petrov et al. (2011).

First, we analyze to which extent system-directed utterances share properties with human-directed utterances. Second, we aim at identifying salient features of intuitively spoken user input. Third, we analyze the impact of the users' age and gender on their speaking style to gain additional insights into the variability of user in-

[2] https://github.com/explosion/spaCy.

put. Therefore, system-directed utterances are compared with human-directed utterances broken down by the users' age and gender. The collected data are examined in terms of different linguistic criteria commonly used in the literature, e.g. (Summa et al., 2016; Johansson, 2008; Pinter et al., 2016; Pak and Paroubek, 2010), including those mentioned by the literature for naturalness:

- Lexical diversity
- Lexical density
- Big words
- POS tag frequencies
- Politeness
- Filler words
- Syntactic complexity
- Sentence types
- Utterance length

Only those features which occur significantly often in system-directed speech are considered as characteristic features of intuitively spoken user input. In order to determine the linguistic features that are associated with the respective criterion, e.g., what is polite, we rely on the findings from literature (see below).

One of the most common measures of lexical diversity is the type-token ratio which is defined as the ratio of the total number of individual word types (lemmas) to the total number of occuring word tokens, cf. (Johansson, 2008). We use the standardized type-token tatio (STTR), firstly mentioned by Johnson (1944), to normalize the impact of the size of the different corpora. Fig. 3 displays the STTR broken down by different age, gender and interlocutor.

The type-token ratio significantly differs between human-directed speech and system-directed speech (p<0.01). In addition, Fig. 3 shows that the older the users the higher the lexical diversity. That is, older participants tend to use more individual words than younger both in system-directed speech and in human-directed speech. The differences between the age groups are significant at p<0.01. The users' gender does not seem to have an impact on the lexical diversity.

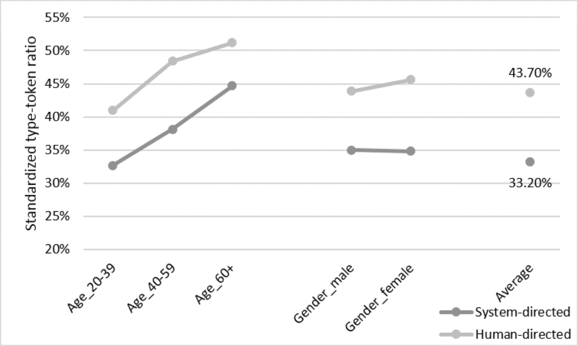

Figure 3: Type-token ratio broken down by user profiles and interlocutor

One of the measures of lexical density is the content-function word ratio which is calculated by dividing the number of content words (open class words) by the number of function words (closed class words), cf. (Johansson, 2008). This means, the higher the proportion of content words the more information is given. In human-directed speech people tend to use more content words (44.68%) than in system-directed speech (41.68%). The user profiles do not seem to have an impact.

The big word ratio is calculated by dividing the number of words longer than six characters (big words) by the total number of words. We found that people do not tend to adapt the use of big words significantly to their interlocutor. 17.11% of the system-directed words are big words and 16.50% of the human-directed. The user profiles do not seem to have an impact on the use of big words.

Next, we are interested in a difference of tag distributions between the speech sets. Table 1 shows the seven most frequent POS tags of both speech sets. Nouns (NOUN) and proper nouns (PROPN) occur much more frequently in the system-directed speech set, whereas adverbs (ADV) and verbs (VERB) occur much more frequently in the human-directed speech set. Pronouns (PRON) are less frequently used in system-directed speech (5.50%) than in human-directed speech (10.42%). The proportion of prepositions (ADP) is ranked at position seven in human-directed speech but at position four in system-directed speech. The proportions of determiners (DET) are more or less balanced. As for the user groups in both sets, we found differences in the occurrence of verbs between men and women. Women tend to use more verbs than men (in system-directed speech significant at the 0.05 level). Additionally, we found that older users tend to use more verbs and pronouns and fewer

Table 1: POS tag frequencies

System-directed		Human-directed	
NOUN	18.93%	ADV	16.73%
PROPN	17.40%	NOUN	14.16%
DET	13.28%	VERB	13.05%
ADP	12.65%	PROPN	12.47%
ADV	12.38%	DET	11.02%
VERB	9.50%	PRON	10.42%
PRON	5.50%	ADP	10.06%

proper nouns than younger people. These tendencies hold for both system-directed speech and human-directed speech.

Our evaluation of how polite users speak to an SDS is based on the empirical findings of (Danescu-Niculescu-Mizil et al., 2013). They characterized politeness marking in requests. Out of the 14 strategies which are perceived as being polite the following strategies appear in our data:

- Sentence-medial please: Could you **please**
- Counterfactual modal: **Could/Would** you...
- Indicative modal: **Can/Will** you...
- 1st person start: **I** search...
- 1st person pl.: Could **we** find...

The distribution of utterances with politeness indicators are shown in Fig. 4.[3] The results in Fig. 4 confirm that politeness strategies are salient features of human-directed utterances but not of system-directed utterances. Overall, only 19.63% of the system-directed utterances contain politeness markers, whereas 53.33% of the human-directed utterances are polite ($p<0.01$). Fig. 4 shows that politeness strategies have been used more often by women in both corpora ($p<0.01$). Furthermore, younger people (20-39 years) are far more likely to avoid politeness strategies when speaking to the system than older people ($p<0.01$).

As for the categorie *filler words*, we investigate the number of utterances that contain disfluencies such as *äh* and *ähm* (eng. "uh") and modal particles. We use the definition of modal

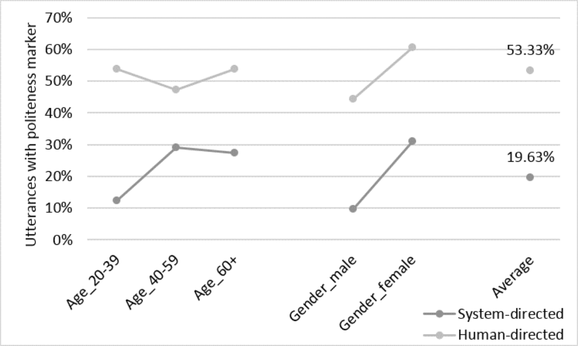

Figure 4: Distribution of polite utterances broken down by user profiles and interlocutor

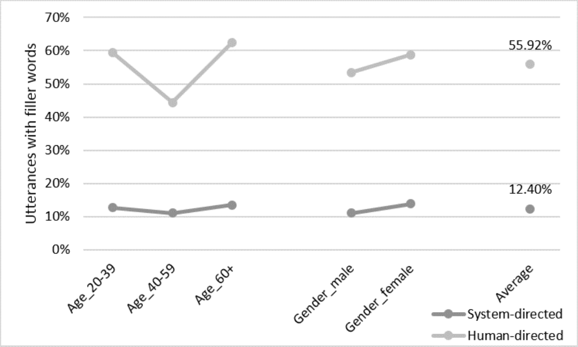

Figure 5: Distribution of utterances containing filler words broken down by user profiles and interlocutor

particles according to Bross (2012), namely that modal particles do not contribute to the sentence meaning. The following modal particles occur in our data: *doch*, *einmal*, *nochmal*, *mal*, *denn*, *eigentlich*, *vielleicht*. Fig. 5 shows the percentage of utterances with disfluencies and modal particles. The results show that all user groups avoid filler words when speaking to the system. Only 12.40% of the system-directed utterances contain filler words. In contrast, 55.92% of the human-directed utterances contain filler words. Significant differences ($p<0.01$) also appear in the use of filler words between the different age groups. 40-59 years old people tend to use less filler words than the younger (20-39) and older (60+) when speaking to their passenger.

Besides lexical and pragmatic aspects we analyze our data in terms of syntactic features. One of the measures of syntactic complexity is tree depth. Tree depth is defined as the number of edges in the longest path from the root node to a leaf, cf. (Pinter et al., 2016). We have calculated the median and mean depth of the dependency

[3] Direct questions such as *What is your native language?*, direct variants such as imperatives and sentence-initial *please* are perceived as being impolite, cf. (Danescu-Niculescu-Mizil et al., 2013). In our data, 8% of all utterances contain an imperative with sentence-medial *please*. Since imperatives with *please* are perceived as not being polite we did not count *please* in this morphosyntactic context.

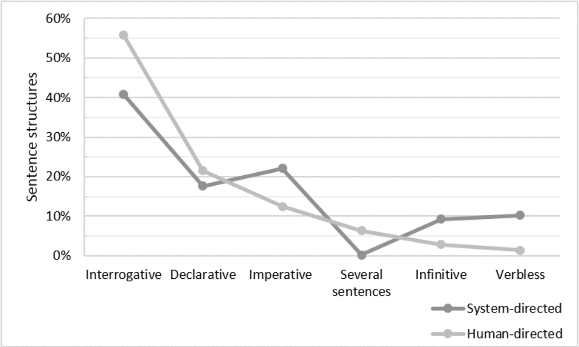

Figure 6: Distribution of sentence structures broken down by interlocutor

trees. However, the differences are not significant at p<0.05. Overall, the median tree depth of the system-directed utterances is 3 with an interquartile range of 2. The same holds for the human-directed utterances.

Another syntactic criterion mentioned by the literature for naturalness is the use of full sentences. The criterion *full sentence* comprises sentences containing a finite verb form. We further subdivided the category *full sentence* into four categories based on sentence types. In addition, we identified patterns without verb or just with an infinitive. We also found utterances composed of two or three sentences that are categorized as *several sentences*. An overview and examples of the sentence structures we identified are given in Table 2. The frequency of the occurrence of the sentence structures is shown in Fig. 6. Across all tasks, an interrogative structure predominates. This is due to the fact that the twelve tasks consist of six information seeking tasks. As Fig. 6 implies, 95,93% of the human-directed utterances are full sentences but only 80,56% of the system-directed. The frequency of an imperative, infinitive and verbless construction increases significantly (p<0.05) in system-directed speech. In human-directed speech people tend to use more interrogative constructions and several sentences to verbalize their request.

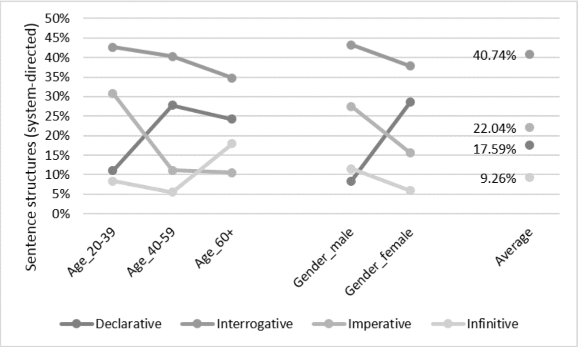

Figure 7: Distribution of sentence structures broken down by user profiles (system-directed)

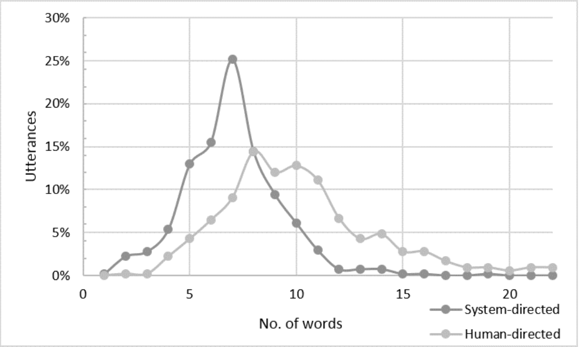

Figure 8: Utterance length broken down by interlocutor

Fig. 7 displays the distribution of sentence structures broken down by user profiles for the system-directed utterances. Only those sentence structures are displayed which show significantly different distributions at the 0.05 level. Younger people (20-39 years) and males tend to use a lot more imperative constructions than older people and females but less declarative constructions. The group of people older than 60 years used more often an infinitive construction than the younger but fewer interrogative constructions. The older participants used fewer interrogative constructions also when speaking to the passenger. As for the distribution of the other sentence structures occured in the human-directed set, the user groups are more or less balanced.

In order to conclude the syntactic analysis we compare the utterance length. Fig. 8 shows the distribution of the number of words per utterance. The utterances towards the system were shorter, Ø 7.01 words per utterance (SD 1.95), than the utterances towards the passenger, Ø 10.22 (SD 3.64).

5 Discussion

Our comparative study shows that certain features, e.g., full sentences or filler words, are characteristic features of interpersonal speaking but not of system-directed speech. We found that although people are told to utter freely they still use syntactic incomplete sentences and they are likely to avoid politeness strategies and filler words, cf. examples given in a) and b).

Table 2: Sentence structures

Sentence Structure	Example
Interrogative	Wo ist die nächste Shell-Tankstelle? “Where is the nearest Shell gas station?”
Imperative	Spiele SWR3! “Play SWR3!”
Declarative	Ich möchte SWR3 hören. “I would like to listen to SWR3.”
Infinitive	SWR3 spielen. No corresponding syntax existing in English
Verbless	Radio SWR3 “Radio SWR3”
Several sentences	Wir könnten ja heute Abend ins Kino. Was kommt denn heute in Berlin? “We could go to the cinema this evening. What's the program in Berlin?”

a) Bitte Radiosender SWR3 einstellen.
“Please radio station SWR3 *infinite verb*”

b) Temperatur auf 23 Grad.
“Temperature to 23 degrees.”

Our analysis results confirm that people adapt their speaking style depending on whom they are talking to. According to the findings of (Levin et al., 2013; Pearson et al., 2006; Branigan et al., 2011) we assume that speakers are strongly influenced by the assumptions that a speaker has about his interlocutor, not only in human-machine communication but also in human-human communication. Thus, people always utter in a way they believe the system is able to understand, also if the system behaves more human-like. We therefore argue that freely spoken user input should not be considered synonymous with human-directed speech, namely with full sentences, civility, with the occurrence of filler words etc. The use of short and concise phrases (such as a verbless construction) just seems to be an effect of the user adapting to the system as conversational partner in the sense of (Pearson et al., 2006; Branigan et al., 2011) and is as natural (in the sense of intuitive) as using full sentences including politeness markers or filler words. If system developers follow the assumption that the linguistics of freely spoken user input is equated with interpersonal speaking style they hardly meet the user expectations of an intuitive and natural speaking. Instead, we suggest to add incomplete syntactic structures such as verbless and infinite sentences to the criteria for naturally spoken user input. Since 71% of the system-directed utterances do not contain filler words or politeness markers we also suggest not to equate natural language input with the occurrence of filler words and politeness indicators.

6 Conclusion

In this paper, we have contributed to the question of how we can define *naturalness* in user input towards a state-of-the-art SDS.

We have presented a user study in which we have collected freely spoken user utterances for a wide range of automotive use cases in German. By means of a comparative study of human-directed and system-directed utterances, we have shown that *naturalness* cannot simply be equated with *human-human* communication: users will use shorter and concise phrases in order to interact with the machine. We have argued that this is an effect of the user adapting to the machine as conversational partner in the sense of (Pearson et al., 2006; Branigan et al., 2011). In addition, we found that the users’ age and gender have an impact on the way they speak to an SDS. We have shown that women did more often make use of politeness strategies and of a declarative construction and that older users tended to use more individual words.

Our further goal is to define evaluation criteria which consider freely spoken user input to compare different SDS. This will be subject of future work.

References

Linda Bell. 2003. *Linguistic Adaptions in Spoken Human-Computer Dialogues: Empirical Studies of User Behavior*. Ph.D. thesis, KTH Royal Institute of Technology.

Markus Berg. 2012. Survey on spoken dialogue systems: User expectations regarding style and usability. In *14th International PhD Workshop OWD*.

Markus Berg. 2013. Natürlichsprachlichkeit in dialogsystemen. *Informatik-Spektrum* 36(4):371–381.

Markus Berg, Petra Gröber, and Martina Weicht. 2010. User study: Talking to computers. In *Proceedings of the 3rd Workshop on Inclusive eLearning*.

Niels O. Bernsen, Hans Dybkjaer, and Laila Dybkjaer. 1998. *Designing Interactive Speech Systems: From First Ideas to User Testing*. Springer.

Peter Bohlin, Johan Bos, Staffan Larsson, Ian Lewin, Collin Matheson, and David Milward. 1999. Survey of existing interactive systems. trindi project deliverable d1.3. Technical report, University of Gothenburg.

Francesca Bonin, Ronald Böck, Nick Campbell, and Ronald Poppe, editors. 2015. *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*. Springer.

Holly P. Branigan and Jamie Pearson. 2006. Alignment in human-computer interaction. In *Proceedings of the Workshop on How People Talk to Computers, Robots, and Other Artificial Communication Partners*.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42:2355–2368.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, Janet F. McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121(1):41–57.

Patricia Braunger, Hansjörg Hofmann, Steffen Werner, and Maria Schmidt. 2016. A comparative analysis of crowdsourced natural language corpora for spoken dialog systems. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

Fabian Bross. 2012. German modal particles and the common ground. *Helikon. A Multidisciplinary Online Journal* 2:182–209.

Nils Dahlbaeck, Arne Joensson, and Lars Ahrenberg. 1993. Wizard of oz-studies – why and how. In *Proceedings of the Workshop on Intelligent User Interfaces*.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Christine Doran, John Aberdeen, Laurie Damianos, and Lynette Hirschman. 2001. Comparing several aspects of human-computer and human-human dialogues. In *Proceedings of the 2nd SigDial Workshop on Discourse and Dialogue*.

Jens Edlund, Joakim Gustafson, Matthias Heldner, and Anna Hjalmarsson. 2008. Towards human-like spoken dialog systems. *Speech Communication* 50:630–645.

Ido Guy. 2016. Searching by talking: Analysis of voice queries on mobile web search. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Akira Hayakawa, Luz Saturnino, and Nick Campbell. 2016. Talking to a system and talking to a human: A study from a speech-to-speech, machine translation mediated map task. In *Proceedings of INTERSPEECH*.

Hansjörg Hofmann, Ute Ehrlich, André Berton, and Wolfgang Minker. 2012. Speech interaction with the internet - a user study. In *Proceedings of the 8th International Conference on Intelligent Environments*.

Hansjörg Hofmann, Mario Hermanutz, Vanessa Tobisch, Ute Ehrlich, André Berton, and Wolfgang Minker. 2014. Evaluation of in-car sds notification concepts for incoming proactive events. In *Proceedings of 5th International Workshop on Spoken Dialog Systems (IWSDS)*.

Victoria Johansson. 2008. *The Department of Linguistics and Phonetics: Working Papers 53*, Lund University, chapter Lexical diversity and lexical density in speech and writing: a developmental perspective, pages 61–79.

Wendell Johnson. 1944. Studies in language behavior: I.a program of research. *Psychological Monographs* 56:1–15.

Scott R. Klemmer, Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, and Annie Wang. 2000. Suede: A wizard of oz prototyping tool of speech user interfaces. In *Proceedings of the 13th Annual ACM Symposium on User interface Software and Technology*.

Daniel T. Levin, Stephen S. Killingsworth, Megan M. Saylor, Stephen M. Gordon, and Kazuhiko Kawamura. 2013. Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human-Computer Interaction* 28(2):161–191.

Sebastian Möller, Florian Gödde, and Maria Wolters. 2008. Corpus aanalysis of spoken smart-home interactions with older users. In *Proceeding of the 6th International Conference on Language Resources and Evaluation (LREC)*.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*.

Bo Pang and Ravi Kumar. 2011. Search in the lost sense of "query": Question formulation in web search queries and its temporal changes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*.

Jamie Pearson, Jiang Hu, Holly P. Branigan, Martin J. Pickering, and Clifford I. Nass. 2006. Adaptive language behavior in hci: How expectations and beliefs about a system affect users' word choice. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI)*.

Slav Petrov, Dipanjan Das, and Ryan T. McDonald. 2011. A universal part-of-speech tagset. *CoRR* .

Yuval Pinter, Roi Reichart, and Idan Szpektor. 2016. Syntactic parsing of web queries with question intent. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Gabriel Skantze. 2007. *Error Handling in Spoken Dialogue Systems: Managing Uncertainty, Grounding and Miscommunication*. Ph.D. thesis, KTH Stockholm.

Anja Summa, Bernd Resch, and Michael Strube. 2016. Microblog emotion classification by computing similarity in text, time, and space. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*.

Masahiko Tateishi, Katsushi Asami, Ichiro Akahori, Scott Judy, Yasunari Obuchi, Teruko Mitamura, Eric Nyberg, and Nobuo Hataoka. 2005. *A Spoken Dialog Corpus for Car Telematics Services*, Springer, chapter DSP for In-Vehicle and Mobile Systems, pages 47–64.

Ryen White, Matthew Richardson, and Wen tau Yih. 2014. Questions vs. queries in informational search tasks. Technical report, Microsoft.

Ute Winter, Tim J. Grost, and Omer Tsimhoni. 2010. Language pattern analysis for automotive natural language speech application. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*.