IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2022, accepted May 4, 2022, date of publication May 12, 2022, date of current version May 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174592

Preferences for Single-Turn vs. Multiturn Voice Dialogs in Automotive Use Cases–Results of an Interactive User Survey in Germany

PETER BURGGRÄF[1,2,3], MORITZ BEYER[1], JAN-PHILIP GANSER[2], TOBIAS ADLON[1], KATHARINA MÜLLER[1], CONSTANTIN RIESS[2], KASPAR ZOLLNER[2], TILL SAßMANNSHAUSEN[3], AND VINCENT KAMMERER[1]

[1]Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, 52074 Aachen, Germany
[2]Management Consultancy StreetScooter Research GmbH, 52066 Aachen, Germany
[3]Department for Mechanical Engineering, University of Siegen, 57076 Siegen, Germany

Corresponding author: Moritz Beyer (m.beyer@wzl.rwth-aachen.de)

ABSTRACT Voice assistants have manifested their existence in the vehicle over the last decade. Further technological developments in the area of voice recognition and interactions with users open up the opportunity for new customer-centric user scenarios. In the following work, eight dialog use cases and two different interaction types were examined in detail. The focus of this work is to answer the research question to which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations. Employing a three-step online survey in 2020, participants were asked about their preferences for the assistant's interaction type, use case as well as the perceived usefulness and pleasantness. The authors found that users preferred multi-turn conversations over single-turn conversations for all defined use case scenarios. As further challenges for the future development of voice assistants in the automotive context, the changed driving situation due to e.g. progress in autonomous driving and the focus on an integration of the voice modality as a direct function should be considered.

INDEX TERMS Automotive, chatbot, voice control, voice communication, voice dialog.

I. MOTIVATION, PROBLEM AND APPROACH

In recent years, voice assistants such as Siri or Alexa have reached an ever-increasing market [1]. The smart speaker has established itself in society through its use at home, in the areas of entertainment, e.g. searching and listening to music, communicating information about the current weather or smart home control functionalities [1]. However, voice assistants have also become increasingly popular in vehicles. Dominant user scenarios are found in the pre-call support, navigation and finding the right radio station [1]. In the future, we might also see voice commands as the standard control modality for certain vehicle functions such as windows opening, closing the doors or controlling the heating from inside or outside the vehicle.

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

According to Grice [2], the successful implementation of a voice dialog between conversation partners depends on the four maxims of quality, quantity, relevance and clarity. A voice dialog can be subdivided within a three-part content classification: dialogs in specific knowledge domains (dialog about e.g. vehicle functions), target-oriented dialogs (dialog to the navigation target), and dialogs with input processing and response generation (e.g. open-domain dialogs about general topics) [3]. A further classification can be made into single-turn and multi-turn conversations. Here, a turn is a cycle of a statement by interlocutor A and a statement by interlocutor B [4]. A single-turn conversation consists of all information provided by one interlocutor in one go and without further conversational turns, e.g. interlocutor A: ''Will it rain tomorrow?'' with the answer from interlocutor B: ''No''. In multi-turn conversations the interlocutors' responses take previous conversational turns in perspective. The conversation information is provided in multiple turns,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

e.g. interlocutor A: ''Will it rain tomorrow?''; interlocutor B: ''No, it will be sunny. What are your plans?''; interlocutor A: ''Maybe we can have a picnic and enjoy the sunny weather?''; interlocutor B: ''Yes, definitely''.

Past research used the above-mentioned maxims and classification variants to locate and evaluate voice assistant developments from a technical point of view [5], [6] to show the rapid progress in the areas of Natural Language Understanding (NLU) and Natural Language Processing (NLP). In this context, even with the latest technical developments, voice assistants cannot accurately predict how the user responds. Funk *et al.* [7] address the gap of different customer responses while identifying an intersection where their developed technology shows intelligence to predict the right customer response and interact accordingly. Further work within the stream of voice assistant research is also focused on developing voice criteria matrices to validate the quality of varying interactions on different platforms [8] and identify and confirm different customer satisfaction for human-machine dialogs. In their study with one user and four defined voice characters (Friend, Admirer, Aunt and Butler), Braun *et al.* [9] identified the increased preference for voice assistants that properly correspond to the personality of the user. However, within automotive use cases, the research is rare. To the authors' knowledge, the work from Braun *et al.* [9] and Funk *et al.* [7] are the two main publications in this field.

Based on the conversation maxims, classification variants, and current research landscape it is not clear which type of dialog is most interesting to the user in a specific in-car situation and how the dialog then should be conducted to optimize user satisfaction. The motivation of this work is thus to answer those questions and give customer-based guidance for the future development of voice control use cases in the automotive industry. In this vein, the authors introduce a study that analyzes the context-specific behavior of users interacting with voice assistants in the vehicle. Based on a voice-interactive online questionnaire, 180 participants took part in the survey and evaluated specific automotive-focused human-machine voice interactions. Thus, this work aims to answer the research question ''*to which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations?*''

The following article is structured into six sections. Section I introduces the current voice assistant research. Section II defines the general communication process, voice dialog systems and gives an overview of the latest voice dialog system development for vehicles. Section III introduces the method of the interactive user survey. Based on this survey, the results will be presented in section IV and discussed in section V. In section VI the findings are summarized and further suggestions and implications for the development of voice assistants in the automotive context are given.

II. COMMUNICATION, VOICE DIALOG SYSTEMS AND CHATBOT DEVELOPMENT

In a communication process information, ideas, facts, or opinions are transmitted between corresponding parties. Laswell [10] described this process in his 1948 publication with the following question: ''Who says what in which channel to whom with what effect?'' Shannon [11] developed a basic communication model in 1948 within the framework of communication research around the same time. Their encoder-decoder model consists of a transmitter (encoder), a message, a channel (voice), a source of interference (noise or interference), and a receiver (decoder) [11]. In his interactive model, which was revised in 2003, DeVito [12] optimized the previous linear models and included missing aspects such as the reciprocal communication structure and feedback. Common to all models, verbal communication contributes one of the central means of dialog in the communication process.

In a dialog, the interlocutors exchange information utilizing statements [13]. A distinction is made between listening, questioning, explaining, laughing, and humoring [14]. A change of the speaker defines the scope of a statement. A cycle of a statement by interlocutor A and interlocutor B is called a turn [4]. The interlocutors can also take on the roles of listener and speaker. This concept of turn-taking leads to an interaction between the dialog partners [15] taking into account the personal intentions and preferences of the interlocutors as well as contextual factors that make each dialog individual and unique.

The focus of this work is on human-machine dialogs. Human-machine interaction refers to the communication and interaction between human users and a computer or a machine exclusively via a human-machine interface [16]. The interaction can be visual, acoustic, or tactile [17]. Human-computer communication can be carried out with different input modules, e.g. a keyboard, a touch screen, or voice recognition software [18]. Those human-machine systems that communicate utilizing voice are referred to as voice dialog systems. They allow to receive spoken utterances as input from the user and to generate utterances as output. Voice dialog systems require several components to work together to recognize the user's voice input, extract the relevant information from the input, retrieve information from back-end services, decide the next step in the dialog and finally generate system responses and synthesis [19]. The initiative to open the human-machine dialog can be taken by the user, by the system (proactivity), or, depending on the use case, adaptive by the user or the system [20].

Such voice dialog systems continue to form the basis for modern chatbot systems that operate with written text or spoken language. Chatbots are classified as systems that can interact with users on a conversational level [19]. They are further categorized into task-oriented and non-task-oriented. Task-oriented systems are designed to perform specific tasks and usually operate in a closed-domain environment e.g. a website's FAQs [21]. Non-task-oriented systems do not

have a specific goal, but enable so-called chit-chat, i.e. conversations with the user in an open-domain environment that cover a general range of topics [22]. Chatbots can be further differentiated into query-based and generative models depending on the technology of input processing and response generation. Query-based models use databases with predefined responses and select a suitable response heuristically based on the input and the context. The complexity of the generation ranges from a rule-based to a machine learning approach. Most studies on query-based systems are upon single-turn question-and-answer conversations [23], which do not generate new statements, but select predefined answers from a finite library and are therefore only effective in closed-domain applications. Retrieval-based systems usually have a lower complexity than generative models, are easier to control and deliver high-quality answers for popular or specific topics. Due to the technological development in conversation-based artificial intelligence (AI), multi-turn question-and-answer systems are increasingly used, which can store and contextualize previous utterances with the help of neural networks [24].

However, concerning the existing literature [7] for voice dialogs in the automotive context, the user preferences are not researched comprehensively, if multi-turn is preferred over single-turn dialogs for certain use cases. Here, the use of voice dialog systems faces different conditions than at home or with the use of a smartphone: A user sits inside the vehicle and cannot leave the environment of the voice assistant while driving. Currently, the driver must always keep an eye on the traffic and cannot use extensive touch navigation or view content on a screen. Voice and audio are the primary input and output means that can be used safely while driving. [25]

Since 2000, in-car voice assistants have been helping to make services available more easily, comfortably and safely while driving. The first produced vehicle with native voice recognition software was the Jaguar S-Type 2000 with voice technology from Dragon Systems UK [26]. These early systems recognized only a limited vocabulary, so users had to adapt to the system's language. With the development of Siri in 2010 and other intelligent personal assistants, the technology has been rolled out for mass production in smart devices. In parallel, vehicle manufacturers pursued the development of their own native assistants. For example, at the CES in 2018, Mercedes-Benz presented its new AI-supported voice assistant. With MBUX (Mercedes-Benz User Experience), users can change vehicle functions via voice input. The statement "I'm cold" leads to an increase in the temperature at the thermostat. The voice control system is capable of learning and adapting to the users and their voices. In addition, it constantly learns new words and grammar, so that the use of language adapts [27]. Furthermore, BMW introduced its AI-supported voice dialog system in 2018, the so-called BMW "Intelligent Personal Assistant". In addition to executing voice instructions, the user can also conduct short conversations on topics such as "Hey BMW, what's the meaning of life?" The assistant can remember preferred settings and proactively suggests changes over time [28].

Cerence [29], a Nuance Communication Inc. spin-off, has the largest customer base for voice systems in the automotive sector. With Cerence Drive, the company provides a robust, automotive-grade portfolio of products, services, toolkits, and innovations, as well as a development platform. Cerence's technology platform enables vehicle manufacturers to let the intelligent assistants interact with the user while the processing of the data takes place in the background by a third-party provider. Alexa Auto, Google Automotive Services, Apple CarPlay, and Android Auto offer similar services as Cerence, with the difference of integrating either user interfaces, or the voice, or avatars into the vehicle from the respective third-party provider. Thus, the user no longer interacts with the vehicle platform itself, but with the offered service in the vehicle. Many vehicle manufacturers implement parallel systems where either e.g. Apple CarPlay or the native assistant can be used. Hybrid solutions recognize the user's intention and determine which of several assistants can best fulfill the requirement and then lead the interaction with the user. [30]–[32]

In summary, the rapid development and optimization of voice dialog systems applied in the automotive context can be noticed on the market side. It aims to provide intelligent, effective, and efficient end-to-end usability, which the user should perceive as high-quality.

Within this matter, the quality of the voice dialog is no longer a mere discussion about the performance in the Turing test or the ability of the system to mimic human-like conversation [33].[1] The quality of voice dialogs nowadays consists of a multitude of attributes that gain relevance in different phases of the system's life cycle depending on the implementation and especially the use case of its application [35]. According to Möller [20], there are individual factors for voice dialog systems that sum up to the resulting perceived quality of the dialogs for the users. Möller [20] distinguishes between generic environmental factors, task factors, system factors, and background factors. At this juncture, it is necessary to examine in greater detail if and to what extent personal dialog preferences, as an additional quality factor for voice dialog systems for vehicles, also contribute to the quality perception of the voice dialog in automotive use cases. A comprehensive summary of the main contributions to the formulation of the implied research question and survey design for this work based on the existing literature is presented in Table 1.

To the authors' knowledge, the consideration of dialog type preferences (multi-turn vs. single-turn) is not yet considered

[1]The test is named after the English mathematician Alan Turing, a pioneer in the field of artificial intelligence. He developed the Turing test concept in 1950. Here, a computer mimics human responses and a study participant is asked to determine several times, whether the answer to a posed question was given by a computer or another human. If the person cannot determine a significant difference between the human or the computer answer, the computer has "passed" the Turing test [34].

TABLE 1. Literature findings with regard to the formulation of the research question and survey design.

Contribution to the formulation of the research question and survey design	Detailed contribution	Sources
Fundamentals of voice assistant design	Basic research on the communication process	[10, 11, 12]
	Basic dialog structure and types of dialogs	[13, 14, 15]
	Definition of interaction modes with human-machine interfaces	[16, 17, 18]
	Natural Language Understanding (NLU) & Processing (NLP)	[19]
Use case design for voice assistants	Proactive voice assistant behavior	[20]
	Specification of task-oriented dialogs	[21]
	Specification of non-task oriented dialogs	[22]
	Specification of single-turn dialogs	[23]
	Specification of multi-turn dialogs	[24]
Voice assistants in the automotive industry	Application context of native voice assistants in vehicles	[25, 26, 27, 28]
	Third party developments of voice assistants for vehicles	[29, 30, 31, 32]
Quality perception of voice assistants	General quality perception of voice assistant dialogs	[33, 34, 35]
	Individual measurement of the perceived quality of voice assistants	[20]

in the literature within automotive use cases explicitly. What makes the automotive use cases especially interesting for further investigation is the fixed position of the user, coupled with a compelling activity focus on the vehicle operation activity (e.g. driving, shifting gears, etc.) and a specific ambient noise (e.g. noise and conversations in the cabin). Here, deviating findings are expected regarding the preferences of the customers when using the voice assistant compared to stationary systems (e.g. smart home or smart device voice assistants).

Therefore, an interactive quantitative study was conducted to investigate the extent to which different personal preferences of the dialog use cases exist for the application area ''automotive'' as well as for preferences of the dialog types (multi-turn vs. single-turn) and causes leading users to their preferences. This will help to increase the quality-perception of automotive voice use cases as well as the decision for specific dialog conducting types. It will guide future development in the area of automotive voice dialogs to focus on user-centric alignment.

III. METHOD: INTERACTIVE USER SURVEY AND USE CASE DESIGN

Section III A) shows the procedure for setting up different types of dialogs for relevant use cases in the application area ''automotive''. An interactive questionnaire with certain dialogs was developed to collect data for this study. Section III B) describes the structure of the questionnaire and the execution of the study.

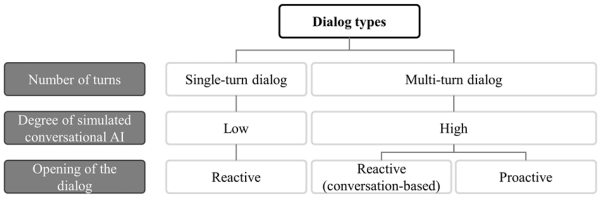

FIGURE 1. Classification of dialog types considered in the customer study.

A. SETTING UP DIALOGS ON USE CASES FOR THE APPLICATION AREA ''AUTOMOTIVE''

Two dialog types have been defined on which the dialogs for the subsequent survey were built. The main distinction was made between the number of turns, the degree of conversational AI simulated, the size of the subject area to be mapped, the degree of system knowledge, and the dialog opening. The differentiating factors for setting up the use cases are based on existing literature in the context of voice assistants [20] and have been validated with expert developers from a premium vehicle manufacturer (OEM) in Germany.

Here, conversational AI in general is the technology making conversations and interactions more human-like; an increase in this field can lead to the development of a personality of the speech applied in human-machine conversations. The AI was simulated using scripted intelligent questions and answers. Concerning the designed use cases, the single-turn dialog was used to answer clearly defined questions with precise answers. The dialog was to be opened reactively, by the user. The system was not capable of learning and could answer a maximum of one question on a specific topic with low to no conversational AI simulated in the dialogs. The multi-turn dialog on the other side was designed task-oriented and gave suggestions for action or information hints for tasks, with an increasing degree of simulated conversational AI. The multi-turn dialog was further divided into a proactive one, where the chatbot starts the conversation with the user and a reactive-conversation-based one, where the user starts the conversation and the chatbot tries to keep the conversation going. For the proactive dialogs, the degree of proactivity is not changed during the interaction with the survey participants. The proactive behavior is limited only to the initiation of the conversation. Nevertheless, the degree of proactivity contributes an important factor for the further chatbot development. With regard to driving a vehicle and operating a proactive chatbot, especially the identification of critical situations, where a proactive opening could have a negative impact on the customer experience should be considered. For the survey in this work, the information was given in advance to the participants that no critical driving situation (e.g. high

velocity, high traffic volume) is present and the vehicle would only show proactive behavior in this case.

Both systems were simulated to show learning behavior and the capability of remembering information about users and their habits. To the authors' knowledge, no such multi-turn dialog systems have been implemented in vehicle cockpits with this degree of conversational AI. Fig. 1 presents an overview of the dialog types considered in the study of this research work.

Based on an expert workshop eight thematic fields for the survey were designed (see Fig. 2). The workshop was conducted with developers from a premium vehicle manufacturer (OEM) in Germany on topics of voice assistants in the driver's cockpit as well as existing use cases of the leading native voice assistants. The thematic fields represent a use case development direction for voice assistants in the automotive industry with regard to enhanced context information, better usability and joy of use for the customer. The resulting thematic fields are *navigation, parking situation, vehicle function, gas station and refueling services, ambient environment, intelligent task manager, vehicle and brand knowledge, and small talk.*

For each thematic field, explicit use case scenarios were set up that represent a specific in-car situation (see Fig. 2). The scenarios are the fastest route, free parking spaces, adaptive cruise control (ACC), filler cap & petrol station services, point of interest (Wilhelmshöhe mountain park), tire status and change, information on specific design details, and others such as e.g. chitchat. The individual use case scenarios are presented in Fig. 2 and merely depict a specific situation from many of the respective superordinate topic areas.

One dialog was developed for each combination of dialog types (1. question and answer dialog; 2. proactive, task-oriented dialog; 3. conversation-based, task-oriented dialog) and for each of the eight use case scenarios (see Fig. 2). The dialogs between a human and the voice assistant were scripted and digitalized in two separate steps. The human sections of the dialog were recorded. The sections of the voice wizard were developed using a text-to-speech (TTS) API from Google. Code was written in Microsoft Visual Studio using the C# programming language, see Appendix I. A female voice speaking in German was used as the assistant's voice, following a majority of companies that use and develop voice assistants. Female voices seem more cordial [36] and are said to be pleasant and sympathetic [37] according to an Amazon representative. A WaveNet voice was chosen as the voice form. This type of synthetic voice sounds more natural than previous voice output systems due to its human-like emphasis [38]. Each dialog section was converted into an MP3 file using the TTS bot and merged with the human recording to form a complete dialog for the study participants.

B. EXECUTION OF THE STUDY AND STRUCTURE OF THE QUESTIONNAIRE

At the beginning of the questionnaire an introduction, introductory questions and a presentation of a voice dialog system

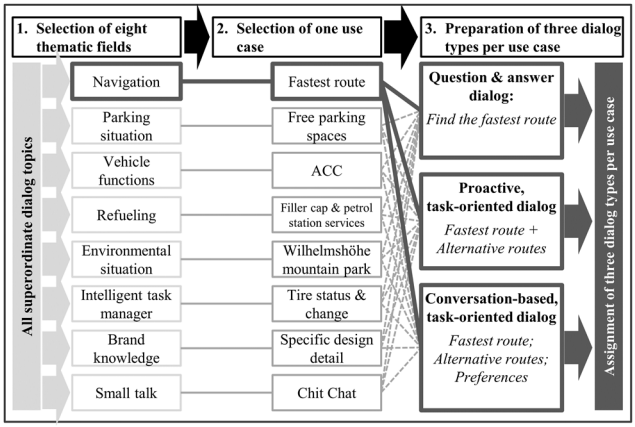

FIGURE 2. Individual use case scenarios with the example for the use case "fastest route."

were given. The main part of the online questionnaire was structured in three coherent sections. In the introductory questions of the questionnaire four self-assessments were recorded, to the participant's (1) technology and (2) automotive affinity as well as the frequency of use of (3) voice control systems in general and (4) voice control in vehicles, using a 5-stage Likert scale. Smart home devices such as (Amazon Alexa) were given as an example of voice control systems in general. Voice control in vehicles referred to voice controls integrated in a vehicle, e.g., for navigation or mobile device control. In the first part, the user had to select three preferred topics from the eight superordinate subject areas to individually reduce the study query to a reasonable number of questions and thus guarantee constant high concentration while answering the questions. The second part asked the user to rate the dialogs for the three corresponding use cases (see Fig. 3). Before moving to the next use case, study participants had to finalize the active use case. For each use case, the three recorded dialogs were played and the participant had to rate the interactions on a 5-stage Likert scale. The dialogs were arranged randomly in each online questionnaire and there was no predefined dialog-play or evaluation order within one preferred topic to avoid framing of the participants. In doing so two questions were evaluated: "*How helpful do you imagine the dialog in your vehicle to be in everyday life?*" and "*How pleasant did you find the dialog*?" In addition, the user had to choose one dialog within one use case as the preferred in the situation. An excerpt for the use case "free parking space" is given in Fig. 3. The questionnaire ends with sociodemographic questions, including gender and age, and a "Thank You" note.

IV. STUDY RESULTS

A. DESCRIPTIVE STATISTICS

The target group of the study was potential customers of automobile manufacturers as well as people with an affinity for vehicles and technology in German-speaking countries. The questionnaire was advertised with recurring postings on

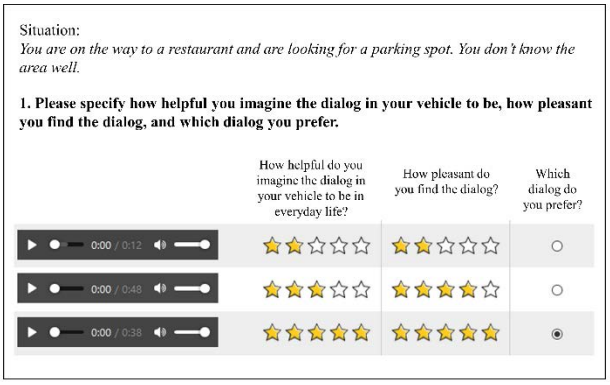

FIGURE 3. Excerpt from the survey on the use case "Parking Situation."

the web platform "linkedin.com" and was online for a total of 18 days in spring 2020. The survey website was viewed approximately 5,000 times, of which a net of 180 people completed the questionnaire in full. This corresponds to a response rate of around 4 %.

The first step of the preparation was to check the total processing time to exclude possible false data sets from the analysis. Participants had to listen and rank different dialogs during the survey. A minimal time could be set in which a participant had time to at least listen to all the dialogs for an appropriate amount of time. The minimum processing time was set at four minutes (240 seconds). This excluded two data sets from the analysis in advance (the total processing time was 109 and 113 seconds respectively). The arithmetic mean of the 178 complete questionnaires was 680 seconds (11:20 minutes) and the median was 12 seconds longer (692 seconds $\hat{=}$ 11:32 minutes). The number of valid questionnaires thus amounted to N = 178. The data were then checked for plausibility to avoid errors in the data processing. The persons' free text entries were adapted so that they corresponded to the generic specifications and could be evaluated. When checking and cleaning the raw data from the listed errors, care was taken not to falsify the data and thus generate false statements. The average age of the participants was M = 33 years with a standard deviation of SD = 12 years. The median was 28 years with a range R = 49 years (age_{max} = 67 years; age_{min} = 18 years). Additionally, the participants were clustered by their age into five groups of ten years, starting with the first group at age 18 to 27, the second from age 28 to 37 and continuing until the fifth age group from age 58 to 67. 61 % of the respondents were male, 39 % female. Out of the 178 participants 51 (28.65 %) drove a vehicle from a premium vehicle manufacturer (e.g.: BMW, Audi, Mercedes-Benz, etc.). The four participants' self-assessments of their (1) technology and (2) automotive affinity and frequency of use of (3) voice control systems in general and (4) voice control in vehicles divide the data sets into two groups: low and high affinity and usage. Depending on the rating of the 5-stage Likert scale, low affinity and usage consisted of ratings of 1, 2 and 3 and high affinity and usage of ratings 4 and 5 on the 5-stage Likert scale. This data is interpreted as interval scaled data. Out of the 178 participants, 117 (65.73 %) have a high technology and 136 (76.04 %) have a high automotive affinity. In total 43 (24.16 %) participants often use voice control in vehicles as well as 29 (16.30 %) participants often use voice control systems in general (in other environments than vehicles). According to the Shapiro-Wilk test, neither of the data above are normally distributed ($p < 0.05$), therefore only nonparametric tests can be used.

In the study, the multi-turn dialogs were divided into proactive- and conversation-based dialogs. This was done due to their different approach of opening a conversation (proactive vs. reactive) and the resulting different influence on the users because both types are a new form of human-machine communication probably not known to the users yet. Hence, these two multi-turn dialogs were chosen to give the user a better understanding of these prospective dialog types. Considering the initial research question ("*To which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations?*") an analysis of the preferred best dialog conducting type is carried out on a higher level: single-turn vs. multi-turn dialogs (see Table 2). Therefore, the proactive and conversation-based dialogs – due to their high degree of congruence of characteristics regarding future-oriented conversation AI – are summarized as "multi-turn dialogs" for the following analysis. The question and answer dialogs are referred to as "single-turn dialogs". A spotlight on differences between the two multi-turn dialog types (proactive vs. conversation-based) is brought up in section V of this work. The descriptive analysis of the study results is shown in Table 2.

B. CORRELATION ANALYSIS OF USER GROUP CRITERIA AND USE CASES

Based on the analysis of Table 2, the majority of the participants chose multi-turn dialogs. However, across all use cases, there is a wide variation between the preferred choices for multi-turn dialogs, ranging from 93.18 % (Intelligent Task Manager) to 56.72 % (Vehicle Functions). Thus, to validate the underlying results on a statistical basis, the following user groups were correlated with the choice of the best dialog type for each use case: gender (female; male), age group (5 groups á 10 years), premium vehicle user (yes; no), technology affinity (low; high), automotive affinity (low; high), use of voice control systems in general (rare; often) and use of voice control in vehicles (rare; often). Based on the user groups introduced above, a total of 48 relationships were analyzed (six user groups with eight use cases) concerning significant correlations (e.g. whether there is a significant relationship between gender and navigation, gender and parking situation, gender and vehicle functions, gender and ambient environment in the choice of multi- or single-turn dialogs as the preferred dialog type). Here, a chi-squared test or Fisher's exact test [39] was used to investigate underlying correlations between the choice of the favorite dialog type per use case and

TABLE 2. Choice of single-turn versus multi-turn dialogs depending on use cases and users.

Topic area	Total number of users N_{tot}[-]	Preference for single-turn dialog		Preference for multi-turn dialog	
		Percentage [%]	Number of users N_s [-]	Percentage [%]	Number of users N_m [-]
Navigation	141	16.31	23	83.69	118
Parking Situation	111	18.02	20	81.98	91
Vehicle Function	67	43.28	29	56.72	38
Ambient Environment	53	30.19	16	69.81	37
Gas Station and Refueling	52	7.69	4	92.31	48
Intelligent Task Manager	44	6.82	3	93.18	41
Vehicle and Brand Knowledge	36	38.89	14	61.11	22
Small Talk	30	33.33	10	66.67	20

the user group, e.g. the gender of a study participant in the use case ambient environment.

Table 3 summarizes all significant correlations within the survey use cases for the analyzed relationships (significance level $p < 0.05$). Detailed further analysis between the six user groups and eight use cases can be found in Appendix II to Appendix VIII. A determination of whether the chi-squared or Fisher's exact test was used depends on the expected cell count number in each contingency table. The expected cell count indicates the frequency that would be expected on average in a matrix cell if the variables were independent. Each 2×2 contingency table consists of four (or ten for a 5×2 contingency table) cells indicating the frequency of correlation.

In the statistical analysis, the asymptotic significance (2-sided) of the chi-squared test was used for cells that have an expected count of five or more [40]. For example, the significant correlation between gender and ambient environment from Table 3 has no expected cells within its 2×2 matrix with a count of five or less.

For cells with an expected cell count less than five, the significance (2-sided) of Fisher's exact test was used [39]. For example, the significant correlation between age group and navigation from Table 3 has three expected cells (30 %) within its 2×5 matrix with a count of five or less.

For cases with a significant correlation, the normalized contingency coefficient C was then determined. Depending on the correlation coefficient C, the strength of the correlation between the variables can be determined: a weak correlation exists for a $C < 0.2$, a medium correlation for $0.2 < C < 0.6$ and a strong correlation for $C > 0.6$ [41].

The overall analysis shows that within 48 relationships of user groups and use cases, 42 (87.5 %) relationships turn out to be non-significant compared to six significant relationships (12.5 %). Within these six significant correlations, the majority decided in favor of the multi-turn compared to the single-turn dialog. All significant correlations show a medium correlation strength. No weak or strong correlation strength can be observed.

TABLE 3. Table of all found significant correlations within the survey use cases.

Correlation between:	Gender [Female; Male]	Premium Vehicle Manufacturer (e.g.: Audi, BMW, …) [Yes; No]	Age Group [1 (18-27); 2 (28-37); 3 (38-47); 4 (48-57); 5 (58-67)]	Technology Affinity [Low; High]	Automotive Affinity [Low; High]	In-car Voice Control Use [Rare; Often]
And:	Ambient Environment	Brand Knowledge	Navigation	Ambient Environment	Gas Station and Refueling	Vehicle Functions
Expected cells with a count less than 5	none	1 (25.00 %)	3(30.00 %)	none	2(50.00 %)	none
Asymptotic significance (2-sided)	0.006	-	-	0.003	-	0.045
Fisher's exact test: Exact significance (2-sided)	-	-	0.043	-	0.014	-
Contingency coefficient C	0.353	0.385	0.257	0.376	0.403	0.237
Correlation strength	medium	medium	medium	medium	medium	medium

C. DEEP-DIVE FOR SIGNIFICANT CORRELATIONS WITHIN USER GROUP CRITERIA AND USE CASES

Based on the statistical analysis, the six significant correlations will be presented in more detail in the following section with decreasing degree of preference for multi-turn dialogs.

1) CORRELATION BETWEEN THE USER GROUP CRITERION "AUTOMOTIVE AFFINITY" AND THE USE CASE "GAS STATION AND REFUELING SERVICES"

The correlation "automotive affinity" & "gas station and refueling services" can be observed in Table 4 for 52 participants and showed in total the highest preference for multi-turn dialogs (48 of 52 participants; 92 %). 42 of 43 participants which select a "high automotive affinity" voted in majority for the multi-turn "gas station and refueling services" dialog (98 %) compared to 6 of 9 participants which selected "low affinity" and voted for the multi-turn dialog with 67 % (see Table 4).

Here, the high degree of participants that describe themselves as automotive-affine and prefer the multi-turn dialog for refueling the car use cases may be explained best with the higher degree of information content (e.g. driving range, information about fuel additives, nearest fueling station, etc.) that is transported with the multi-turn dialogs rather than with the single-turn dialog (e.g. only information where the filler cap can be found is provided).

2) CORRELATION BETWEEN THE USER GROUP CRITERION "AGE GROUP" AND THE USE CASE "NAVIGATION"

The correlation "age group" & "navigation" shows the second-largest preference for multi-turn dialogs within

TABLE 4. Overview of the descriptive statistics for the correlation between the user group criterion "automotive affinity" and the use case "gas station and refueling services."

n= 52 (100 %)		Use case: Gas station and refueling services	
		Single-turn	Multi-turn
User group criterion: Automotive affinity	High automotive affinity	1 (2 %)	42 (98 %)
	Low automotive affinity	3 (33 %)	6 (67 %)
	∑	4 (8 %)	48 (92 %)

TABLE 5. Overview of the descriptive statistics for the correlation between the user group criterion "age group" and the use case "navigation."

n= 141 (100 %)		Use case: Navigation	
		Single-turn	Multi-turn
User group criterion: Age group	18-27	7 (11 %)	59 (89 %)
	28-37	5 (14 %)	32 (86 %)
	38-47	2 (16 %)	11 (84 %)
	48-57	6 (43 %)	8 (57 %)
	58-67	3 (28 %)	8 (72 %)
	∑	23 (17 %)	118 (83 %)

TABLE 6. Overview of the descriptive statistics for the correlation between the user group criterion "gender" and the use case "ambient environment."

n= 53 (100 %)		Use case: Ambient environment	
		Single-turn	Multi-turn
User group criterion: Gender	Male	5 (17 %)	25 (83 %)
	Female	12 (53 %)	11 (47 %)
	∑	17 (33 %)	36 (67 %)

"navigation" use cases. It can be found in Table 5 that 118 (83 %) from a total of 141 participants voted in favor of multi-turn dialogs. Within the five defined age groups, the majority always voted in favor of the multi-turn dialog. Within the first age group (18-27 years), 59 of 66 participants (89%) preferred the multi-turn dialog. Followed by the second group (28-37 years), where 32 of 37 participants (86%) preferred the multi-turn dialog. Within the third group (38-47 years), 11 of 13 participants preferred the multi-turn dialog (84%). Within the fourth group (48-57 years), 8 of 14 participants preferred multi-turn dialogs (57 %). For participants within the age group five (58-67 years), 8 of 11 preferred multi-turn dialogs (72 %) (see Table 5).

As it can be expected for this correlation, a significant decline for the preference of multi-turn dialogs within the navigation control from younger to older participants can be observed. The strong preference of younger participants (e.g. 18-37) for multi-turn dialogs within navigation use cases may be attributed to the popular use of voice assistants within today's smart-device navigation functionalities e.g. Apple's Siri – which has naturalized especially into the daily life of younger participants. Interestingly, an incline for the preference of multi-turn dialogs can be observed for the "oldest" age group (72 %) compared to the second "oldest" age group (57 %). This observation will be investigated in future research work.

3) CORRELATION BETWEEN THE USER GROUP CRITERION "GENDER" AND THE USE CASE "AMBIENT ENVIRONMENT"

The correlation between "gender" & "ambient environment" can be observed for 53 participants in Table 6. Here 36 participants (67 %) prefer the multi-turn dialog. Among the female voters, 11 of 23 voted in favor of the multi-turn dialog (47 %). In comparison, 25 of 30 male participants (83 %) voted in favor of the multi-turn dialog (see Table 6).

Here, the authors designed the use case example "Wilhelmshöhe mountain park", where detailed information regarding an ambient sight is given. The "neutral" use case was set up to provoke the same interest among men and women. Nevertheless, it cannot be neglected fully that there is a certain preference within men or women for it – which might explain the underlying difference in the preference for multi-turn or single-turn dialogs. Thus, further research regarding different preferences for certain ambient environment use cases accounting for the gender will be required.

4) CORRELATION BETWEEN THE USER GROUP CRITERION "TECHNOLOGY AFFINITY" AND THE USE CASE "AMBIENT ENVIRONMENT"

The correlation "technology affinity" & "ambient environment" can be observed in Table 7 for 53 participants with a favor of the multi-turn dialog from 36 participants (67 %). 22 participants chose a "low technical affinity" and of those 12 (54 %) voted in favor of the single-turn dialog. However, 31 selected a "high technical affinity" and of those 26 (83 %) voted in favor of the multi-turn dialog (see Table 7).

As it can be expected for this correlation, with a higher technological affinity a higher preference for multi-turn dialogs within ambient environment use cases (Wilhelmshöhe mountain park) can be observed. The strong preference of technology-affine participants for multi-turn dialogs within this use case may also be attributed to the expected popular use of voice assistants within today's smart-device & app functionalities - which is used for example as a guide when exploring new sights or spending time in unknown locations.

TABLE 7. Overview of the descriptive statistics for the correlation between the user group criterion "technology affinity" and the use case "ambient environment."

n= 53 (100 %)		Use case: Ambient environment	
		Single-turn	Multi-turn
User group criterion: Technology affinity	Low technology affinity	12 (54 %)	10 (46 %)
	High technology affinity	5 (17 %)	26 (83 %)
	∑	17 (33 %)	36 (67 %)

TABLE 8. Overview of the descriptive statistics for the correlation between the user group criterion "premium vehicle manufacturer" and the use case "brand knowledge."

n= 37 (100 %)		Use case: Brand knowledge	
		Single-turn	Multi-turn
User group criterion: Premium vehicle manufacturer	No premium car	14 (53 %)	12 (47 %)
	Premium car	1 (9 %)	10 (91 %)
	∑	15 (41 %)	22 (59 %)

TABLE 9. Overview of the descriptive statistics for the correlation between the user group criterion "in-car voice control use" and the use case "vehicle functions."

n= 68 (100 %)		Use case: Vehicle functions	
		Single-turn	Multi-turn
User group criterion: In-car voice control	Low usage	26 (47 %)	27 (53 %)
	High usage	3 (20 %)	12 (80 %)
	∑	29 (41 %)	39 (59 %)

5) CORRELATION BETWEEN THE USER GROUP CRITERION "PREMIUM VEHICLE MANUFACTURER" AND THE USE CASE "BRAND KNOWLEDGE"

The correlation "premium vehicle manufacturer & brand knowledge" can be observed in Table 8 for 37 participants and showed in total a multi-turn preference for 22 participants (59 %). Among the 26 participants which selected "no premium car", 14 (53 %) preferred the single-turn dialog. In comparison, 11 participants selected "premium car". Here, 10 (91 %) participants voted in favor of the multi-turn dialog (see Table 8).

The above analysis should be considered with caution due to the overall low number of participants. Nevertheless, a strong preference of participants, who own a premium vehicle, for multi-turn dialogs within the use case brand knowledge (e.g. explain details about the car and manufacturer history) can be observed. It might be explained, that further brand knowledge will not contribute a major point of interest for participants, who do not own a vehicle from a premium brand. This would align with the preference for multi-turn dialogs since more information can be transported than with single-turn dialogs.

6) CORRELATION BETWEEN THE USER GROUP CRITERION "IN-CAR VOICE CONTROL USE" AND THE USE CASE "VEHICLE FUNCTIONS"

The correlation between "in-car voice control use" & "vehicle functions" can be observed for 68 participants in Table 9. 39 participants (57 %) voted in favor of the multi-turn dialog. Among the 53 participants which selected "low usage [of in-car voice control]", 27 (53 %) preferred the multi-turn dialog. In comparison, within the group of "high usage [of in-car voice control]", 12 of 15 participants (80 %) selected the multi-turn dialog (see Table 9).

For participants, which use in-car voice control more often, especially the vehicle function control via multi-turn dialog contributes to the preferred choice of dialog types. In the authors' opinion, if vehicle users are experienced with the advantages of in-car voice control within other fields of usage (e.g. navigation, multi-media control, etc.), they are generally more open to applying the voice-control to another function field as well. Especially for information regarding complex vehicle functions (Adaptive Cruise Control), voice control offers a great possibility to give a comprehensible introduction – which could explain the observed correlation.

To answer the research question "*To which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations?*" especially the scope of *"certain driving situations"* was elaborated further in this work. Therefore, use cases were set in correlation with user groups. Within the correlations of use cases and user groups for in-car voice assistants, 6 of 48 relationships (see Table 3) could be identified as significant.

The use cases, "ambient environment", "brand knowledge", "navigation", "gas station and refueling services", and "vehicle functions" show significant correlations with certain user groups as discussed this section. On the other side, the use cases "parking situation", "intelligent task manager", and "small talk" do not show any significant correlations with certain user groups.

Regarding the user groups who are described with specific control variables, "gender", "premium vehicle manufacturer", "age group", "technology affinity", "automotive affinity", and "in-car voice control use" show significant correlations with certain use cases as discussed in this section. Only the "use of voice control systems in general" shows no significant correlation with any use case.

Due to the overall low number of significant correlations (6 from 48 possible correlations), no generic abstraction of preferences for certain use cases shall be undertaken. Concerning Möller's [20] taxonomy, the quality of voice communication systems depends on environmental, system, task and background factors. Following this conceptual thought, each use case can be perceived differently in certain situations from different users (user groups). The authors argue – based on the low number of significant use case correlations within the findings – that rather than focusing on specific use cases (e.g. navigation, chit chat), voice assistants should be capable of dealing with this highly individualized matter, which is specific to each user in a certain situation. Since this work focuses on describing the user perspective for voice assistant requirements, the question of how to obtain the configuration for such an adapting voice dialog system remains a challenge in future work.

D. BOOTSTRAP ANALYSIS

With the evaluation shown above, the research question "*To which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations*" can't be confirmed fully due to the low number of significant correlations. Nevertheless, within the six significant relationships, a clear majority decided in favor of the multi-turn dialogs. To support this hypothesis on a general level, the structures of the data were reconsidered and looked at from a different angle. The data were summarized independently of the user group and use case, hence each rating of a person's use case was considered individually so that a total of 534 cases could be analyzed (three use case ratings per participant, 178 participants in total). Out of the 534 cases, 119 times (22.28 %) the single-turn question and answer and 415 times (77.72 %) the multi-turn dialog was chosen as the favorite dialog.

For further evidence, the initial 5-stage Likert scales of whether the dialog was pleasant or useful were distinguished in terms of multi- or single-turn, but not in terms of use cases. In the case of multi-turn dialogs, this leads to a merging of the ratings into a mean value. Both the 5-stage Likert scales and the choice of the favorite dialog are not normally distributed according to the Shapiro-Wilk test and were interpreted as interval scaled data.

To apply parametric tests to the non-normal distributed data, bootstrapping resampling must be applied [42]. Corresponding to Lunneborg [43], a bivariate correlation was performed using bootstrap to analyze the linear relationships between the ratings of whether the dialog was pleasant or useful and the best-rated dialog type. The results are based on 10,000 bootstrap samples. Within this context, Table 10 shows the significant correlations, using Pearson's r (value range -1 to 1), between the choice of the best dialog and the rating of the usefulness or pleasantness for each dialog type individually. Here, Pearson's r is significant on the $p < 0.01$ level, if the upper and lower confidence interval limits do not intersect with zero. A negative prefix of

TABLE 10. Correlation between the choice of the best dialog and its rating of usefulness or pleasantness using bootstrapping.

Choice of the best dialog	Statistical key figures		Rating of usefulness and pleasantness of dialogs: Users' *usefulness* rating of *single-turn* dialog [5-stage-Likert scale]	Users' *usefulness* rating of *multi-turn* dialog [5-stage-Likert scale]	Users' *pleasantness* rating of *single-turn* dialog [5-stage-Likert scale]	Users' *pleasantness* rating of *multi-turn* dialog [5-stage-Likert scale]
Users choice of the best dialog [0: single-turn; 1: multi-turn]	Pearson's r		-0.367**	0.491**	-0.286**	0.373**
	Significance (2-tailed) p		1.835×10^{-18}	1.098×10^{-33}	1.582×10^{-11}	5.074×10^{-19}
	N		534	534	534	534
	95 % Confidence Interval	Lower	-0.438	0.414	-0.360	0.288
		Upper	-0.290	0.562	-0.209	0.453

**. Correlation is significant at the 0.01 level (2-tailed).

Pearson's r describes a significant relationship in preference of the single-turn dialog; vice versa a positive prefix is a preference of the multi-turn dialog.

With the bootstrap correlation analysis, it can be shown that "useful" and "pleasant" assessments are good predictors for the choice of the favorite dialog type (single- or multi-turn) due to their significant correlations (see Table 10). "Useful" is the stronger predictor than "pleasant" comparing the Pearson's r for each dialog type. A high rating of **useful single-turn dialogs** has a medium significant correlation to the choice of single-turn as the favorite dialog ($r = -0.367$, $p = 1.835 \times 10^{-18}$) and a high rating of **pleasant single-turn dialogs** has a medium significant correlation to the choice of single-turn dialog as the best dialog ($r = -0.286$, $p = 1.582 \times 10^{-11}$). Pearson's $r_{single-turn,useful}$ absolute value is higher than Pearson's $r_{single-turn,\ pleasant}$ absolute value ($|r_{single-turn,useful}| = 0.367 > |r_{single-turn,\ pleasant}| = 0.286$). Vice versa a high rating of **useful multi-turn dialogs** has a strong significant correlation to the choice of multi-turn dialogs as the best dialog ($r = 0.491$, $p = 1.098 \times 10^{-33}$) and a high rating of **pleasant multi -turn dialogs**has only a medium significant correlation to the choice of multi-turn dialogs as the best dialog ($r = 0.373$, $p = 5.074 \times 10^{-19}$). Pearson's $r_{multi-turn,useful}$absolute value is higher than Pearson's $r_{multi-turn,\ pleasant}$absolute value ($|r_{multi-turn,useful}| = 0.491 > |r_{multi-turn,pleasant}| = 0.373$) (see Table 10).

In conclusion, the general preference for multi-turn dialogs over single-turn dialogs can be confirmed. On the one hand, this is based on the descriptive analysis. Multi-turn dialogs are favored by more than three quarters (77.72 %) of the choices for the best dialogs across all use cases. On the other hand, the decision "the choice for the best dialog" correlates with the evaluation of the individual dialog types. Here, the correlations in Table 10 show that users who rate multi-turn dialogs as pleasant and useful also select multi-turn dialogs as the best dialog. By analogy, users who perceive single-turn

dialogs as more pleasant and useful are more likely to select single-turn dialogs as the best dialog type. Within the rating for multi-turn dialogs, ''useful'' ($r = 0.491$) has the stronger linear correlation to best dialog choice as ''pleasantness'' ($r = 0.373$). This finding also applies to single-turn dialogs, where ''useful'' ($r = -0.367$) has a stronger correlation to the best dialog choice as ''pleasantness'' ($r = -0.286$).

In this study, the focus was the potential predictors ''useful'' and ''pleasant''. In case of no dependence, pleasantness and usefulness would have been excluded for the time being. Since no other factors were considered in detail, further studies would be required to show on what predictor the choice of the favorite dialog type (single- vs. multi-turn) depends on - e.g. artificial voice or explicit dialog content.

V. DISCUSSION

The results of the survey show that users generally prefer multi-turn conversations over single-turn dialogs within the vehicle. This finding was pointed out with descriptive statistics and could not be refuted using systematic correlation (contingency coefficient) analysis. As a consequence, the study provides important contributions to the voice communication research and answers the question ''*To which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations?*''

A. CONTRIBUTION TO THE VOICE COMMUNICATION SCIENTIFIC COMMUNITY

Firstly, users prefer multi-turn over single-turn in-car conversation with a voice assistant. From this finding, the authors derive the interest of users to interact with the vehicle assistant more frequently. It can be assumed, that the preference for multi-turn conversations finds its foundation in the voice assistant's technology of smartphones and smart speakers – having naturalized into daily life. However, due to the overall low number of significant correlations between use cases and user groups, no generic abstraction of preferences for certain use cases shall be undertaken as mentioned in section IV.

In the context of multi-turn dialogs, a further distinction can be made between conversation-based and proactive dialogs. The analysis in this work gives a first view of the preference for multi-turn dialogs in general without taking into account this corresponding distinction. In a nutshell, the main difference lies with the proactive start of the conversation triggered from the in-car voice control system based upon certain events or frequent polling. Here, the first tendency in favor of proactive multi-turn dialogs over conversation-based multi-turn dialogs could be observed. Out of the 415 participants preferring multi-turn dialogs, 228 participants prefer proactive multi-turn dialogs (54.94 %) and 187 participants prefer conversation-based multi-turn dialogs (45.06 %). With the research focus of this work on the number of turns preferred in automotive situations, the differentiation between proactive and conversation-based dialogs is not the main research contribution. Nevertheless, detailed information for this preference could represent a major research area are for the future since it will have a great influence on the holistic dialog and voice control system design. In this context, the finding of multi-turn dialogs being preferred over single-turn dialogs in automotive use cases lays the foundation for the next research steps towards the preferences for proactive vs. conversation-based dialogs. As a next step, the authors thus recommend further field investigation within this matter. This could support the finding that multi-turn dialogs are preferred over single-turn dialogs if it can be neglected that the preference for the multi-turn dialogs is only based on the characteristics of the proactive or conversation-based dialog (and not from the number of turns).

In addition, it can be stated that in the context of the evaluation of the best dialog by the study participants, both in multi-turn and single-turn dialogs, the assessment criterion ''usefulness'' shows a stronger correlation with the best dialog than the assessment criterion ''pleasantness''. Therefore, the assumption arises that participants perceive usefulness as a more important factor than the degree of pleasantness in the decision of choosing the best dialog type. This in turn has fundamental implications for the future development of voice communication systems in the automotive context and corresponding study design.

Concluding, this study reveals the general importance of multi-turn conversations in the automotive context. However, as the present study is performed as an online survey, the authors were not able to ensure that participants knew exactly the user scenario of every use case. The next step would be to perform the study within a realistic vehicle environment and under driving conditions. More precisely, users need to sit in a vehicle and communicate directly with the voice assistant to ensure the user's understanding of the use case and the system's performance. Due to the low number of observations for the use cases ''intelligent task manager'', ''vehicle and brand knowledge'' and ''small talk'', the results shall be considered with caution and a retesting with a greater number of observations is recommended.

Furthermore, the differentiating factors (number of turns, degree of conversational AI simulated, size of the subject area to be mapped, degree of system knowledge, dialog opening) for the dialogs in the third section for setting up the initial use cases were derived from literature and validated with expert developers from a premium vehicle manufacturer. Future research should investigate which other customer relevant factors might be taken into account and which are most representative for describing automotive voice assistant use cases.

B. FUTURE CHALLENGES FOR VOICE COMMUNICATION AND CONTROL IN AUTOMOTIVE USE CASES

In the authors' opinion, a voice assistant which can interact according to the users' linguistic preferences and communication behaviors creates the basis for pleasant human-machine communication. This development will enable in-car systems to become a useful and enjoyable

tool in the vehicle of the future. A more user-focused voice communication system should also be perceived from the customer side as a high-quality end-to-end system, which can be assumed the goal of premium vehicle manufacturers. Within this context, the authors present three fields, where further development in the automotive industry is required.

In the automotive context, the voice communication modality has historically been used as an additional operating modality to control underlying vehicle functions. Since then, it is evolving with the increasing functional scope of possible in-car use cases. Especially in the last vehicle generations of the premium manufacturers, the scope of voice communication has increased due to more intelligent and personalized use cases, e.g. in the multi-media and navigation domain. Concerning the vast autonomous driving developments, the authors argue that voice communication will become more important as novel types of vehicle use can be expected – e.g. the vehicle as a working atmosphere, as a family multi-media entertainment park. The upcoming challenge here is to align voice assistants with conversation dialog types (multi-turn vs. single-turn) and future use cases required in the automotive context.

Furthermore, in the thematic fields of this survey, only ''vehicle and brand knowledge'' and ''small talk'' represent direct use cases with a focus on communication between the vehicle and the user. The use of voice communication in the other use cases can rather be seen as an extension of the controllability for existing functions (e.g. navigation, parking situation, vehicle function, gas station and refueling services, ambient environment, and intelligent task manager). In this context, other use cases of voice processing in vehicles will become important in the future, dealing for example with authorization functions via personalized voice recognition. The already mentioned functionality of the MBUX system of the S-Class (2020) [27] can be seen as a leading example of the latest development, which gives the option of voice recognition for authorization to log in to personalized user profiles. As shown in the above example, the development perspective for voice assistants needs to widen its view of taking voice as only an additional operating modality for certain use cases for creating specific use cases around the powerful voice communication and control possibility. The observed general preference of multi-turn use cases for the exchange between the vehicle and the user via voice (see section IV) can be seen as an argument, that this would generally be accepted from the customer side.

Another challenge will be the design of voice assistants in the exterior concerning single or multi-turn dialog guidance as well as relevant use cases. Voice assistants in the exterior would allow the customer to interact with the vehicle even when they are not in the passenger compartment. This could be relevant for the customer when he wants to e.g. open certain doors or the hatch of the vehicle. This novel application possibility represents a barely developed field of research and holds great potential for further exploration. What makes this situation interesting and completely different to inside-vehicle services of today will be for example the interaction outside of the car with pedestrians [44] and the question of what use cases and conversation types are preferred from the customer in general.

VI. CONCLUSION

Voice control of certain tasks and functions is becoming increasingly important in the automotive sector – also driven by the broad acceptance of voice assistants in the home environment and smart devices. In this context, the question arises of what constitutes a suitable dialog in the automotive domain (use case type) and how it should be conducted (multi-turn vs. single-turn). This paper, therefore, focused on the research question ''*to which extent do users prefer task-oriented multi-turn dialogs over the question-and-answer single-turn dialogs in certain driving situations?*'' In this work, an interactive user study was carried out using specially coded voice assistants for certain use cases freely selectable by the survey participants. In total, a net sample of N = 178 could be evaluated as part of the data collection. The evaluation of the results showed that users in the automotive context generally prefer multi-turn over single-turn dialogs, no matter the use case. Concerning use cases development, no specific preferences were derived from the significant correlations found, which is justified by the individuality of each vehicle driver, the driving situation and other contextual factors – as already described in Möller's [20] work. The obtained results from this work and the scope of future research should be acknowledged with regard to the current limitations of the study as mentioned in the discussion section. Future voice assistants need to be developed in the direction of dealing with individual factors in any kind of situation. As further challenges for the future development of voice assistants in the automotive context, the changed driving situation due to e.g. progress in autonomous driving and the focus on an integration of the voice modality as a direct function should be considered. Likewise, the development of voice assistants for the vehicle exterior could contribute to a significant field of research in the future. How to obtain the optimal configuration for those challenges remains future work.

ACKNOWLEDGMENT

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] B. Kinsella and A. Mutchler. (2020). *In-Car Voice Assistant Consumer Adoption Report. Edited by Voicebot.Ai*. Accessed: Apr. 13, 2021. [Online]. Available: https://voicebot.ai/wp-content/uploads/2020/02/in_car_voice_assistant_consumer_adoption_report_2020_voicebot.pdf

[2] P. Grice, *Studies in the Way of Words*, 3rd ed. Cambridge, MA, USA: Harvard Univ. Press, 1993.

[3] Botpress. *What is a Chatbot: Everything You Need to Know in 2022 | Botpress*. Accessed: Apr. 13, 2021. [Online]. Available: https://botpress.com/learn/chatbots

[4] S. W. Hamerich, *Sprachbedienung im Automobil: Teilautomatisierte Entwicklung Benutzerfreundlicher Dialogsysteme*. Berlin, Germany: Springer, 2009.

[5] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces," in *Advances in Intelligent Systems and Computing, Advances in Human Factors and Systems Interaction*, I. L. Nunes, Ed. Cham, Switzerland: Springer, 2018, pp. 241–250, doi: 10.1145/3313831.3376816.
[6] E. V. Polyakov, M. S. Mazhanov, A. Y. Rolich, L. S. Voskov, M. V. Kachalova, and S. V. Polyakov, "Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning," in *Proc. Moscow Workshop Electron. Netw. Technol. (MWENT)*, Moscow, Russia, Mar. 2018, pp. 1–5, doi: 10.1109/MWENT.2018.8337236.
[7] M. Funk, V. Tobisch, and A. Emfield, "Non-verbal auditory input for controlling binary, discrete, and continuous input in automotive user interfaces," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Honolulu, HI, USA, Apr. 2020, pp. 1–13, doi: 10.1145/3313831.3376816.
[8] J. Jiang, A. Hassan Awadallah, R. Jones, U. Ozertem, I. Zitouni, R. Gurunath Kulkarni, and O. Z. Khan, "Automatic online evaluation of intelligent assistants," in *Proc. 24th Int. Conf. World Wide Web*, Florence, Italy, May 2015, pp. 506–516, doi: 10.1145/2736277.2741669.
[9] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging, and F. Alt, "At your service: Designing voice assistant personalities to improve automotive user interfaces," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Scotland, U.K., May 2019, pp. 1–11, doi: 10.1145/3290605.3300270.
[10] H. D. Laswell, "The structure and function of communication in society," in *The Communication of Ideas*, L. Bryson, Ed. New York, NY, USA: Harper and Brothers, 1948, pp. 37–51.
[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jun. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
[12] J. A. DeVito, *The Interpersonal Communication Book*, 14th ed. Boston, MA, USA: Pearson, 2016.
[13] D. Jurafsky and J. H. Martin, "Speech and language processing," in *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 2020.
[14] J. Röhner and A. Schütz, *Mittel Verbaler Kommunikation*, in *Psychologie der Kommunikation*. Wiesbaden, Germany: Springer, 2012, pp. 71–83, doi: 10.1007/978-3-531-18891-1_5.
[15] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, vol. 50, no. 4, pp. 696–735, 1974, doi: 10.1353/lan.1974.0010.
[16] P. Ulrich and W. Hill, "Wissenschaftstheoretische grundlagen der betriebswirtschaftslehre," *Wirtschaftswissenschaftliches Studium: Zeitschrift für Ausbildung und Hochschulkontakt*, vol. 5, nos. 7–8, pp. 304–309, 1976.
[17] GlobalSpec. *Human Machine Interface (HMI) Software Selection Guide: Types, Features, Applications | Engineering360*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.globalspec.com/learnmore/industrial_engineering_software/industrial_controls_software/human_machine_interface_software_hmi
[18] J. K. Petersen, *The Telecommunications Illustrated Dictionary*. Boca Raton, FL, USA: CRC Press, 2018, doi: 10.1201/9781315220222.
[19] M. F. McTear, "The rise of the conversational interface: A new kid on the block?" in *Future and Emerging Trends in Language Technology. Machine Learning and Big Data* (Lecture Notes in Computer Science), J. F. Quesada, F.-J. Martín Mateos, T. López Soto, Eds. Cham, Switzerland: Springer, 2017, pp. 38–49, doi: 10.1007/978-3-319-69365-1_3.
[20] S. Möller, "Quality of transmitted speech for humans and machines," in *Communication Acoustics*, J. Blauert, Ed. Berlin, Germany: Springer-Verlag, 2005, pp. 163–192, doi: 10.1007/b139075.
[21] A. Rudnicky, E. Thayer, P. Constantinides, C. Tchou, R. Shern, and K. Lenzo, "Creating natural dialogs in the Carnegie Mellon communicator system," in *Proc. 6th Eur. Conf. Speech Commun. Technol. (EUROSPEECH)*, Budapest, Hungary, Sep. 1999, pp. 1531–1534. Accessed: Apr. 13, 2021. [Online]. Available: https://www.isca-speech.org/archive_v0/archive_papers/eurospeech_1999/e99_1531.pdf
[22] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The design and implementation of XiaoIce, an empathetic social chatbot," *Comput. Linguistics*, vol. 46, no. 1, pp. 53–93, Mar. 2020, doi: 10.1162/COLI_a_00368.
[23] H. Wang, Z. Lu, H. Li, and E. Chen, "A dataset for research on short-text conversations," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, WA, USA, 2013, pp. 935–945. Accessed: Apr. 13, 2021. [Online]. Available: https://aclanthology.org/D13-1096.pdf
[24] R. Lowe, N. Pow, I. Serban, and J. Pineau, "The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems," in *Proc. 16th Annu. Meeting Special Interest Group Discourse Dialogue*, Prague, Czech Republic, 2015, pp. 285–294, doi: 10.18653/v1/W15-4640.
[25] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: The effect of speech-based E-Mail on drivers' attention to the roadway," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 43, no. 4, pp. 631–640, Dec. 2001, doi: 10.1518/001872001775870340.
[26] M. Krebs. *Behind the Wheel/2000 Jaguar S-Type; A More Compact Cat With a Classic Face*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.nytimes.com/1999/07/25/automobiles/behind-the-wheel-2000-jaguar-s-type-a-more-compact-cat-with-a-classic-face.html
[27] Daimler AG. *Mercedes-Benz MBUX Multimedia System: The New Digital Cockpit*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.mercedes-benz.de/passengercars/technology-innovation/mbux/mbux-stage.module.html
[28] N. Horn and B. Titz. *Hey BMW, Jetzt Reden Wir*. Mit BMWs Intelligent Personal Assistant bekommt der BMW Eine Persönlichkeit. Accessed: Apr. 13, 2021. [Online]. Available: https://www.press.bmwgroup.com/deutschland/article/detail/T0284429DE/%E2%80%9Ehey-bmw-jetzt-reden-wir-%E2%80%9C-mit-bmws-intelligent-personal-assistant-bekommt-der-bmw-eine-persoenlichkeit?language=de
[29] *Cerence, Solutions*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.cerence.com/solutions
[30] *Google LLC, Android Auto & Android*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.android.com/intl/de_de/auto/
[31] Apple. *iOS—CarPlay*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.apple.com/ios/carplay/
[32] *Amazon, Echo Auto—Bringen Sie Alexa in Ihr Auto: Amazon.de: Amazon-Geräte & Zubehör (Engl.: Echo Auto—Bring Alexa to Your Car: Amazon.de: Amazon Devices & Accessories)*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.amazon.de/echo-auto-bringen-sie-alexa-in-ihr-auto/dp/B078YP59TT
[33] A. Törrönen. *Why Chatbots Need Quality Standards?* Accessed: Apr. 13, 2021. [Online]. Available: https://www.chatbotpack.com/chatbot-quality
[34] G. Oppy and D. Dowe. *The Turing Test*. Accessed: Apr. 13, 2021. [Online]. Available: https://plato.stanford.edu/entries/turing-test/#
[35] N. M. Radziwill and M. C. Benton, "Evaluating quality of chatbots and intelligent conversational agents," 2017, *arXiv:1704.04579*.
[36] W. J. Mitchell, C.-C. Ho, H. Patel, and K. F. MacDorman, "Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management," *Comput. Hum. Behav.*, vol. 27, no. 1, pp. 402–412, 2011, Indianapolis, IN, USA, doi: 10.1016/j.chb.2010.09.002.
[37] H. Schwär and Q. Moynihan. *Companies Like Amazon May Give Devices Like Alexa Female Voices to Make Them Seem 'caring'—Business Insider*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.businessinsider.com/theres-psychological-reason-why-amazon-gave-alexa-a-female-voice-2018-9?r=DE&IR=T
[38] *Google LLC, WaveNet und Andere Synthetische Stimmen (Engl.: WaveNet and Other Synthetic Voices)*. Accessed: Apr. 13, 2021. [Online]. Available: https://cloud.google.com/text-to-speech/docs/wavenet
[39] R. A. Fisher, "Statistical methods for research workers," in *Biological Monographs and Manuals*, no. 5, F. A. E. Crew and D. W. Cutler, Ed. Edinburgh, U.K.: Oliver and Boyd, 1954.
[40] D. S. Yates, *The Practice of Statistics: TI-83 Graphing Calculator Enhanced*. New York, NY, USA: W.H. Freeman, 1999.
[41] IBM. *Cramér's V*. Accessed: Apr. 13, 2021. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_ca_dshb.doc/cramersv.html
[42] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, PA, USA: SIAM, 1982, doi: 10.1137/1.9781611970319.
[43] C. E. Lunneborg, "Estimating the correlation coefficient: The bootstrap approach," *Psychol. Bull.*, vol. 98, no. 1, pp. 209–215, 1985, doi: 10.1037/0033-2909.98.1.209.
[44] S. Sadeghian, M. Hassenzahl, and K. Eckoldt, "An exploration of prosocial aspects of communication cues between automated vehicles and pedestrians," in *Proc. 12th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, New York, NY, USA, Sep. 2020, pp. 205–211, doi: 10.1145/3409120.3410657.

PETER BURGGRÄF was born in 1980. He received the M.B.A., Dr.-Ing., and Professor degrees. He studied mechanical engineering in Aachen and London. He wrote his Ph.D. Thesis at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, in the field of factory planning. From 2011 to 2017, he was the Chief Engineer with the Chair of Production Engineering, WZL. Since 2013, he has been the Managing Director of StreetScooter Research GmbH. Since 2017, he has been with the Chair of International Production Engineering and Management (IPEM), University of Siegen, Germany.

MORITZ BEYER was born in 1991. He received the M.Sc. degree. He studied industrial engineering at the Rheinisch-Westfälische Technische Hochschule (RWTH), Aachen, from 2011 to 2015, and at the Technische Universität München (TUM) as well as the University of Economics and Business in Athens, from 2015 to 2018. Since 2019, he has been working as a Research Associate at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University. Since June 2021, he has been the Group Lead of the "Digital Factory Planning" Group. Previously, he worked with StreetScooter Research GmbH as a Consultant in the field of product & production design, from 2015 to 2018.

JAN-PHILIP GANSER was born in 1986. He received the Dipl.-Wirt.-Ing. degree in industrial engineering from the Rheinisch-Westfälische Technische Hochschule Aachen (RWTH Aachen), Aachen, Germany, in 2012. From 2013 to 2015, he was a Research Associate and from 2015 to 2017, he was a Group Lead and a Senior Consultant at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University. Since 2018, he has been the Vice President Automotive of StreetScooter Research GmbH.

TOBIAS ADLON was born in 1989. He received the M.Sc. degree from RWTH. He studied industrial engineering at RWTH Aachen University and Imperial College London, specializing in mechanical engineering and production technology. After working in two management consultancies and the manufacturing industry, primarily in the automotive sector, he joined the Chair of Production Systems, WZL, as a Research Associate, in 2015. From July 2018 to March 2020, he was the Group Lead of the "Assembly Planning" Group. In April 2020, he took over as a Chief Engineer with the Factory Planning Department.

KATHARINA MÜLLER received the M.Sc. degree from RWTH. She studied mechanical engineering with a specialization in production engineering at RWTH Aachen University, from 2009 to 2016. Since 2017, she has been a Research Associate with the Factory Planning Department, Chair of Production Systems (WZL), RWTH Aachen University. The focus of her industrial and research projects in Germany and abroad include factory and assembly planning, supply logistics, and production planning.

CONSTANTIN RIESS was born in 1990. He received the B.Sc. and M.Sc. degrees in management (strategy) and technology (electrical engineering) (TUM-BWL) from the Technical University Munich (TUM), Munich, Germany, in 2018. From 2016 to 2018, he was a Research Associate and a Management Consultant at the Institute for Corporate Management, Logistics and Production, TUM. Since 2018, he has been a Management Consultant at StreetScooter Research GmbH.

KASPAR ZOLLNER was born in 1993. He received the B.Sc. and M.Sc. degrees in mechanical engineering and production engineering from the Rheinisch-Westfälische Technische Hochschule Aachen (RWTH Aachen), Aachen, Germany, in 2020. From 2013 to 2016, he was a Research Associate at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, with a research focus in grinding technologies. Since 2020, he has been Co-Founder of s2k digital GmbH.

TILL SAßMANNSHAUSEN was born in 1990. He received the dual M.Sc. degrees. He is currently pursuing the Ph.D. degree with the Chair of International Production Engineering and Management (IPEM), University of Siegen. He studied production engineering at the University of Siegen and business psychology at the FOM University of Applied Sciences, Siegen. Since 2018, he has been working with the IPEM Chair, where he has been managing the production Management Group, since 2019.

VINCENT KAMMERER was born in 1995. He received the B.Sc. degree in industrial engineering from RWTH Aachen University, Germany, in 2020, where he is currently pursuing the master's degree in industrial engineering. From 2019 to 2021, he was a Student Assistant at the Chair of Production Systems, RWTH Aachen University. His research interests include factory planning, artificial intelligence, and operations research.