Advances in Transportation, Vol. 60, 2022, 508–515
https://doi.org/10.54941/ahfe1002484

(Don't) Talk to Me! Application of the Kano Method for Speech Outputs in Conditionally Automated Driving

Deike Albers[1], Niklas Grabbe[1], Yannick Forster[2], Frederik Naujoks[2], Andreas Keinath[2], and Klaus Bengler[1]

[1]Chair of Ergonomics, Munich School of Engineering and Design, Technical University of Munich, 85748 Garching, Germany
[2]BMW Group, Knorrstr. 147, 80937 Munich, Germany

ABSTRACT

Conditionally automated driving (L3) implies repeated transitions of the driving responsibility between the human operator and the automated driving system. This research examines users' attitudes towards speech outputs as potential features for human-machine interfaces for L3 automated driving. The Kano method is applied to identify scenarios where users prefer speech outputs. After a test drive with an L3 automated vehicle, N = 42 drivers take part in a survey on speech outputs in different scenarios. The results identify users' preferences for speech outputs in critical situations. In non-critical application areas, the findings of the Kano method and further comments show a large variance among participants. Customization is desired for the design of speech outputs - including the way of addressing the human operator. Future research should focus on the development of user preferences for speech outputs in long-term studies and the identification of user groups.

Keywords: Kano method, Speech output, Conditionally automated driving, User survey, Human-machine interface

INTRODUCTION

In March 2021, the first Level 3 automated system was launched in Japan (Sugiura 2021). Conditionally automated driving (Level 3) implies that the driving task is completely handed over to the automated driving system (ADS), while the human operator stays responsive to resume the driving task in cases of ADS-issued requests to intervene or system failures (J3016). This level of automation entails repeated transitions of the responsibility for the driving task between the human operator and the ADS. Human-machine interfaces (HMI) may facilitate these transitions.

To communicate with the driver, HMIs can apply different output channels such as visual, auditory, or haptic channels (Bengler et al. 2020). Combinations of modalities lead to faster reaction times (e.g., Burke et al. 2006) and perceived urgency (e.g., van Erp et al. 2015). For auditory cues, simple tones, complex sounds, and speech outputs may be applied. Speech outputs have the advantage of being gaze-free (Meng und Spence 2015). They have

been shown to reduce visual workload during driver-automation interaction and to convey more information than haptic or visual interfaces (Bazilinskyy und Winter 2015). On the downside, Campbell et al. (2007) and Naujoks et al. (2016) assume that repeated speech outputs might reduce acceptance by users. Most research on speech outputs has been done in the context of ADS-issued requests to intervene. For specific instructions, speech outputs may be used in combination with visual displays (Perez et al. 2009). The authors of UR:BAN (2016) recommend using speech outputs in the early phases of a warning cascade. Regarding users' opinions, combinations of sound, visual, and voice signals receive high ratings (Bazilinskyy und Winter 2015). In an application area different from requests to intervene, Naujoks et al. (2016) were able to show the benefit of speech outputs in situations where the ADS informed the human operator about upcoming maneuvers conducted by the ADS: Users preferred added speech outputs over generic auditory signals and reported reduced visual workload and interference with non-driving related tasks. The findings suggest that speech output might be a beneficial feature in HMIs for L3 automated driving. To the knowledge of the authors, no research has been conducted that focused on the users' preference for speech outputs differentiating between different scenarios, e.g., status information vs. requests to intervene.

The research aim of this paper is to gain detailed insight into the users' perspective on speech outputs for L3 HMIs. The Kano method is applied to allow conclusions to be drawn for speech outputs in different scenarios. The Kano Model as introduced by Kano (1984) shows the relationship between customer satisfaction and the functionality of product features. For the identification and categorization of speech output features for different application areas of L3 HMIs, the Kano method is applied according to Shen (1993).

Though the paper does not focus on the design of speech outputs, the survey includes users' preferences for being addressed in either a personalized manner or a distanced manner. A previous study conducted by Danner et al. (2020) did not find an effect on usability ratings regarding how the human operator was addressed. Furthermore, participants are requested to share their thoughts on speech outputs in general.

METHOD

The experiment is part of a larger study that will be published elsewhere. Only information relevant to this paper is provided here. The overall experimental design is presented in Albers et al. (2021).

Study Design

The study was conducted between July 2021 and August 2021 on a test track at the BMW Driving Academy in Maisach, Germany. A BMW 3 Series (G21) was modified for Level 3 automated driving. A sociodemographic questionnaire was provided remotely prior to the experiment. Upon arrival, participants were instructed on the experimental procedure and gave informed consent. A familiarization drive was followed by a test drive of about 45 minutes. During the test drive, participants experienced twelve use

cases covering continuous rides in automation levels L0, L2, and L3 (J3016), changes in the availability of automation levels, and transitions between automation levels, both initiated by the participant and initiated by the ADS. After the test drive, participants were asked to fill out a questionnaire covering the survey on speech outputs. The overall duration of the experiment was two hours.

Sample

The sample consists of N = 42 drivers ($M = 38.21$, $SD = 9.73$, $n_{female} = 19$, $n_{male} = 23$). In a between-subject design, participants were assigned to two HMI concepts. None of these HMI concepts contained speech outputs. Therefore, no distinction between the participants is made in this paper.

Scenarios & Resulting Speech Output Features

Six features for potential speech outputs are covered by the Kano method. Table 1 presents a short description of each feature. Four of the features refer to scenarios that all participants experienced during the test drive (Albers et al. 2021). The test cases cover changes in the availabilities of automation levels, confirmations of successful transitions, and ADS-issued requests to intervene. Two features refer to situations experienced only by some of the participants: a hands-off warning, and feedback for operating errors. Only participants that showed specific behavior, e.g., taking their hands away from the steering wheel while driving L2, received a notification.

The first part of the survey on speech outputs covered the Kano questionnaire. For each feature two questions were presented (Shen 1993): The first question is the functional form of the pair of questions ("How would you feel if you had this feature/speech output?"). The second question is the dysfunctional form ("How would you feel if you did NOT have this feature/speech output?"). The five possible answers are: 0: "I like it"; 1: "I expect it"; 2: "I am neutral"; 3: "I can tolerate it" and 4: "I dislike it" (Pouliot 1993).

After the Kano questions, participants were asked whether the vehicle should communicate in a passively distanced manner or whether the vehicle should address the driver in an actively personal manner. Examples were provided for passively distanced and actively personal, respectively: "Automated driving now available" and "You can activate automated driving now, Tom", respectively. Participants indicated their preference on a 5-point semantic differential from "definitely passively distanced" to "definitely actively personal". Finally, participants were invited to add comments on speech outputs in a free text box.

RESULTS

Kano Method

First, a discrete analysis is conducted (Shen 1993). For every participant, the answers to individual features are transposed into four categories of the Kano

Table 1. Description and examples of the speech output features covered in the Kano method. Corresponding test cases experienced in the test drive prior to the survey are indicated (as presented in Albers et al. 2021).

Feature	Description	Example	Test case(s)
Availability Change-Up	Automation levels not previously available now become available.	Assisted and automated driving is now available.	2
Availability Change-Down	Previously available automation levels are now no longer available.	There is a sensor error. Automated driving is currently not available.	6
Transition	You switch to another automation level.	Automated driving is now active.	3, 5, 9-12
Request To Intervene	You are told that you will soon have to switch to manual driving.	System limit ahead. Please switch to manual driving soon.	10
Hands-Off Warning	The assisted driving level is active, and you do not have your hands on the steering wheel. You are told to put your hands on the steering wheel.	Please put your hands on the steering wheel.	None; only experienced if corresponding behavior was shown
Operating Error	Manual driving is active. You press the toggle key, which switches between assisted and automated driving. This key has no function when manual driving is active.	Activate assisted driving first to use the automated driving function.	None; only experienced if corresponding behavior was shown

method: *Must-Be*, *Performance* (or *One-Dimensional*), *Attractive*, and *Indifferent*. *Must-Be* features of a product are expected. For *Performance* features, customer satisfaction increases proportionally with functionality. *Attractive* product features are not expected by customers but cause a positive reaction. *Indifferent* product features do not matter to the customer, they may be present or absent. Two further categories indicate whether the individual participants give contradicting answers for specific features (*Questionable*) or answers that suggest that specific features are not wanted (*Reverse*). Each feature is attributed to the category with the highest score. A Fong-Test is applied to test for statistical significance (Fong 1996 as cited in Hölzing 2008). The results are presented in Table 2.

The three features *Availability Change-Down, Transition,* and *Request to Intervene* are assigned to the category *Performance*. The assignment is significant for the features *Availability Change-Down* ($Fong_{left}=12, Fong_{right}=7.13$) and *Request to Intervene* ($Fong_{left}=12$, $Fong_{right}=7.42$). The feature *Transition* does not receive significant results ($Fong_{left}=1$, $Fong_{right}=6.74$). Answers on the feature *Availability Change-Up* result in a non-significant assignment to category *Attractive* ($Fong_{left}=5$, $Fong_{right}=6.91$) and the feature *Hands-Off Warning* is assigned to category *Must-Be* (non-significant:

Table 2. Discrete analysis of the Kano method. The categories are M: *Must-Be*, P: *Performance*, A: *Attractive*, I: *Indifferent*, R: *Reverse*, and Q: *Questionable*. The asterisk indicates statistical significance after the Fong-Test (Fong, 1996 as cited in Hölzing, 2008).

Feature	M [%]	P [%]	A [%]	I [%]	R [%]	Q [%]	Category
Availability Change-Up	10	24	36	21	10	0	A
Availability Change-Down	19	48	12	19	2	0	P *
Transition	12	29	26	24	10	0	P
Request to Intervene	26	55	7	7	5	0	P *
Hands-Off Warning	29	29	7	29	7	0	M
Operating Error	14	17	14	40	14	0	I *

$Fong_{left}$=0, $Fong_{right}$=6.83). The feature *Operating Error* is assigned to the category *Indifferent* (significant: $Fong_{left}$=10, $Fong_{right}$=6.83). The features *Availability Change-Up*, *Transition*, and *Operating Errors* receive answers in the category *Reverse* from at least 10% of the participants. There are no answers assigned to the category *Questionable*.

Second, a continuous analysis in the manner of DuMouchel (1993) is conducted and presented in Figure 1. This analysis puts stronger weight on extreme answers and aggregates participants' scores to a position on a two-dimensional coordinate system. Standard deviations are added to display the variance (one-sided for readability reasons).

All features show a high variance with minimum standard deviations for the feature *Request to Intervene* ($SD_{functional}$=1.39, $SD_{dysfunctional}$=1.48). The features *Request to Intervene* and *Availability Change-Down* show clear allocations to the category *Performance*. The feature *Operating Error* is clearly assigned to the category *Indifferent*. The features *Hands-Off Warning*, *Transition*, and *Availability Change-Up* are in the *Performance* and *Must-Be* (*Availability Change-Up)* quadrants, respectively. However, these features' positions are close to other categories.

Preference for Manner of being Addressed by the Vehicle

Participants' preferences regarding the way they are addressed by an ADS show a high variance among participants. More than half of the participants prefer to be addressed in a passively distanced manner. About every fifth participant indicates a preference for actively personal addressing. The endpoints of the scale are used by 50% of the participants (38.10% "definitely passively distanced"; 11.10% "definitely actively personal"). The question does not distinguish between scenarios such as those presented in the Kano method.

Comments on Speech Outputs

About half of the participants (n = 23) make comments on speech outputs. Five participants express their potential annoyance at speech outputs in general or under specific circumstances, e.g., too many prompts or interruptions to music. Eight participants state that speech outputs are particularly (n = 4) or only (n = 4) relevant in critical situations. One participant adds that speech

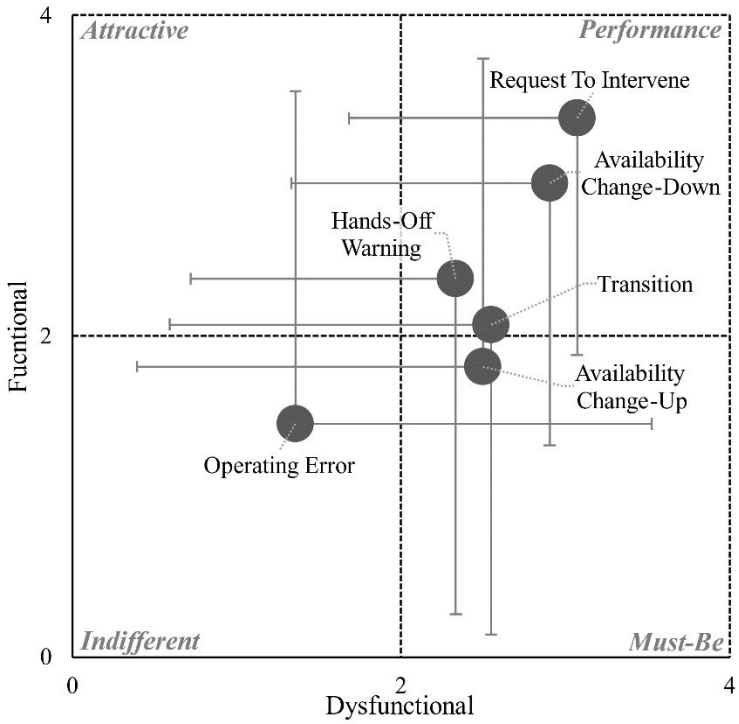

Figure 1: Continuous analysis of the Kano method according to DuMouchel (1993). The six presented features are graphically located in the four categories *Must-Be, Performance, Attractive,* and *Indifferent*. For readability reasons, standard deviations are only presented one-sided.

outputs are helpful for long messages displayed on the dashboard. Nine participants express their wishes on the design of speech outputs. Four of them state that users should be able to turn speech outputs off. Three participants state that the ability to customize is desired in general, for volume, or the voice. The desire for a "nice, calming voice" is expressed by one participant. Two participants elaborate on the manner of being addressed by the ADS: one participant finds "being addressed personally […] creepy"; another participant expects personal addressing to grab the attention of the driver better than passive addressing.

DISCUSSION

The results obtained in the Kano method show large variances among the participants. Both, the discrete and the continuous analysis show clear assignments to categories for only half of the features. The features *Availability Change-Down* and *Request To Intervene* receive clear assignments to the category *Performance*. The example for the feature *Availability Change-Down* does not imply a request for action, but rather information about a sensor error affecting higher (non-active) levels of automation. The feature *Request To Intervene* implies an imminent action by the human operator. In conclusion, both features refer to rather critical scenarios. This is in line with the comments where about 20% of the participants suggested using speech

outputs only in critical situations. The feature *Operating Error* is clearly assigned to the category *Indifferent*. This feature receives as many answers in the category *Must-Be* (14%) as the category *Reverse* (14%). Likewise, the features *Availability Change-Up* and *Transition* receive 10% of the answers in the category *Reverse*. These results indicate that in non-critical application areas, users' attitudes towards speech outputs differ significantly. The user needs may vary depending on their familiarity with the ADS. Speech Outputs for the features *Operating Error, Transition*, and *Availability Change-Up* may be helpful for new users but annoying for experienced users. These results on users' attitudes support the literature findings presented in the introduction stating that speech outputs are useful for critical situations such as requests to intervene (UR:BAN 2016) while frequent speech outputs may lead to annoyance (Naujoks et al. 2016).

Insights gained on the design of speech outputs emphasize the need for customizable solutions for speech outputs where this is possible, e.g., voice settings. Regarding the preferred way of being addressed by the ADS, participants varied widely in their responses. This may be one reason for the findings of Danner et al. (2020) who did not see an effect on usability ratings regarding how the human operator was addressed.

This research offers more detailed insights into users' attitudes towards speech outputs for HMIs for L3 ADS in different scenarios. However, this research is subject to limitations. Though participants experienced driving in L3 mode prior to the survey, speech outputs were only presented as examples in written form. Future research may provide speech output examples that are less abstract, possibly implemented as prototypes. Longitudinal studies, possibly with customizable prototypes could address the development of preferences over time, the identification of different user groups, e.g., with a focus on cultural effects or effects due to prior experiences, and acceptance in naturalistic settings, e.g., with music playing or conversations with passengers.

CONCLUSION

The research presented in this paper identifies users' preferences for speech outputs in critical situations. In non-critical application areas, the findings of the Kano method and further comments show a large variance among participants. These findings suggest the need for customizable speech outputs where this is possible. Customization is likewise desired regarding the design of speech outputs - including the way of addressing the human operator. Future research should focus on the development of user preferences for speech outputs in long-term studies and the identification of user groups.

REFERENCES

Albers, Deike; Radlmayr, Jonas; Grabbe, Niklas; Hergeth, Sebastian; Naujoks, Frederik; Forster, Yannick et al. (2021): Human-Machine Interfaces for Automated Driving: Development of an Experimental Design for Evaluating Usability. In: Nancy L. Black, W. Patrick Neumann und Ian Noy (Hg.): Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021), Bd. 221 (Lecture Notes in Networks and Systems), S. 541–551.

Bazilinskyy, Pavlo; Winter, Joost de (2015): Auditory interfaces in automated driving: an international survey. In: PeerJ Computer Science. DOI: 10.7717/peerj-cs.13.

Bengler, Klaus; Rettenmaier, Michael; Fritz, Nicole; Feierle, Alexander (2020): From HMI to HMIs: Towards an HMI Framework for Automated Driving. In: Information 11 (2), S. 61. DOI: 10.3390/info11020061.

Burke, J.; Prewett, M.; Gray, A.; Yang, L.; Stilson, F.; Coovert, M. et al. (2006): Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In: ICMI '06: Proceedings of the 8th International Conference on Multimodal Interfaces, S. 108–117.

Campbell, J.; Richard, C.; Brown, J.; McCallum, M. (2007): Crash Warning System Interfaces: Human Factors Insights and Lessons Learned. Final Report. DOT HS 810 697. Washington.

Danner, Simon; Pfromm, Matthias; Bengler, Klaus (2020): Does Information on Automated Driving Functions and the Way of Presenting It before Activation Influence Users' Behavior and Perception of the System? In: Information 11 (1), S. 54. DOI: 10.3390/info11010054.

DuMouchel, William (1993): Thoughts on Graphical and Continuous Analysis. In: Center for Quality of Management Journal 2 (4), S. 20–23.

Fong, D. (1996): Using the self-stated importance questionnaire to interpret Kano questionnaire results. In: Center for Quality of Management Journal 5 (3), S. 21–24.

Hölzing, Jörg A. (2008): Die Kano-Theorie der Kundenzufriedenheitsmessung. Eine theoretische und empirische Überprüfung. 1. Aufl.: Gabler.

J3016, 2018: (R) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

Kano, N. (1984): Attractive quality and must-be quality. In: Hinshitsu (Quality, The Journal of Japanese Society for Quality Control) (14), S. 39–48.

Meng, Fanxing; Spence, Charles (2015): Tactile warning signals for in-vehicle systems. In: Accident; analysis and prevention 75, S. 333–346. DOI: 10.1016/j.aap.2014.12.013.

Naujoks, Frederik; Forster, Yannick; Wiedemann, Katharina; Neukum, Alexandra (2016): Speech improves human-automation cooperation in automated driving. In: Mensch und Computer 2016 – Workshopband. Unter Mitarbeit von B. Weyers und A. Dittmar. Aachen.

Perez, M. A.; Neale, V. L.; Kiefer, R. J. (2009): Test and Evaluation of the Cooperative Intersection Collision Avoidance System for Violations (CICAS-V) Driver Vehicle Interface. In: 21st Int'l Tech Conf Enha Safety Vehi (ESV), S. 1–15.

Pouliot, Fred (1993): Theoretical Issues of Kano's Methods. In: Center for Quality of Management Journal 2 (4), S. 28–36.

Shen, Diane (1993): Developing and Administering Kano Questionnaires. In: Center for Quality of Management Journal 2 (4), S. 7–12.

Sugiura, Eri (2021): Honda launches world's first level 3 self-driving car. Plan for just 100 lease sales reflects Japanese automaker's cautious approach. Nikkei Asia. Online verfügbar unter https://asia.nikkei.com/Business/Automobiles/Honda-launches-world-s-first-level-3-self-driving-car, zuletzt geprüft am 24.11.2021.

UR:BAN (2016): 4. Meilensteinbericht (Q16) – Evaluation der funktionsspezifischen MMI Konzepte in den Funktionsträgern und Demonstratoren. Teilprojekt Stadtgerechte Mensch- Maschine-Interaktion - MMI (2016).

van Erp, Jan B.F.; Toet, Alexander; Janssen, Joris B. (2015): Uni-, bi- and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. In: Safety Science 72, S. 1–8. DOI: 10.1016/j.ssci.2014.07.022.