"Hey, Siri"; "Ok, Google"; "Alexa". Acceptance-relevant factors of virtual voice-assistants

Laura Burbach
RWTH Aachen University
Aachen, Germany
burbach@comm.rwth-aachen.de

Patrick Halbach
RWTH Aachen University
Aachen, Germany
halbach@comm.rwth-aachen.de

Nils Plettenberg
RWTH Aachen University
Aachen, Germany
plettenberg@comm.rwth-aachen.de

Johannes Nakayama
RWTH Aachen University
Aachen, Germany
nakayama@comm.rwth-aachen.de

Martina Ziefle
RWTH Aachen University
Aachen, Germany
ziefle@comm.rwth-aachen.de

André Calero Valdez
RWTH Aachen University
Aachen, Germany
calero-valdez@comm.rwth-aachen.de

***Abstract*—Today, virtual voice-assistants are used for manifold purposes. Besides their promising potential, many users are concerned about their privacy and what happens to the data recorded by the voice assistant. It is not yet clear which factors contribute to the acceptance of voice assistants. To address this issue, we conducted a Choice-Based-Conjoint Analysis with three attributes and three levels each. Relating the well-known privacy-utility trade-off, we found out that not the price of a voice assistant is the most important factor for its acceptance, but privacy. Nevertheless the acceptance of voice assistance and the decision to use a voice assistant always depends on a combination of different factors, of which privacy seems to be most important. Besides, four different potential target groups of virtual voice assistants with different preferences could be identified. In the future, the user should be placed in the focus of research, since different configurations are probably desirable for each user.**

***Index Terms*—Voice assistants, virtual assistants, technology acceptance, conjoint study, privacy**

I. Introduction

Virtual voice assistants are one of the major changes in user interaction and user experience design in the recent past. They already allow for many tasks like asking for information, turning off the lights, playing music and the like and are still learning with every interaction users make. This intuitive way of interacting with technical devices without the need for haptic contact [1], [2] makes verbal communication the new interface to technology.

In 2015, [3] found out that more than half (55%) of all U.S. teens use voice search every day and that most teens (89%) and adults (85%) agree when asked whether voice search is going to be 'very common' in the future. It can be expected, that VAs will be used even more in the future due to two developments: First quality of speech recognition will further increase [4] because broadband internet [5] allows for more complex data processing in powerful data centers and each request provides further training to the algorithms. Second, from the users' perspective, VAs facilitate interaction [2].

In the long term, these systems are intended to automate repetitive tasks in companies, for example. Among other things, Amazon's Alexa can open the video conferencing system and enter the corresponding dial-in data or book meeting rooms. Voice assistants will probably be able to truly and fully understand the users at some point. But, do users really want that? In addition to technological progress, the attitudes and concerns of (potential) users will also play a major role in the coming years.

As shown, most people suspect that virtual voice assistants will be increasingly used in the future. Despite the general willingness to use voice assistants in the future, many individuals do not want to use them yet or only for certain occasions. So far, very little research has been done into which aspects contribute to the acceptance or rejection of voice assistants. Further, previous studies usually looked at the evaluation of single factors in regard to the acceptance of a specific voice assistant. Nevertheless, they didn't consider multiple possible product configurations or different user requirements as well as trade-offs between the individual factors. To examine these trade-offs and potential target groups, Conjoint analyses, which are a method to find out user preferences, are very suitable [6]–[8]. So far, to the best of our knowledge, no conjoint studies on the acceptance of voice assistants have been conducted. Therefore, in this study we consider how three specific aspects influence the acceptance of voice assistants and how these three aspects, *NLP-performance*, *price* and *privacy* weighed against each other to examine individual trade-offs and potential user segments of virtual voice assistants.

II. Related Work

Virtual voice assistants (VAs) are a new technology whose acceptance, as already mentioned, has not been researched extensively, yet. Nevertheless, there are some studies using basic technology acceptance models as well as a few studies that have already investigated the acceptance of VAs.

A. Current use of voice assistants

Individuals integrate VAs into their everyday office or everyday life: “Hey, Siri. Write an e-mail to my colleague Mr. Smith.”, “Hi, Cortana. What are my appointments today?” or “Ok, Google , call my friend Alexandra.”

Although voice assistants have great potential to support their users in their actions, their use is currently still limited. At the moment, they are mainly used to call people, ask for directions or search the internet for information. One reason for the limited use can be, that users are dissatisfied if errors occur at the automated speech recognition. Interactions between users and virtual assistants, such as prompting the device to call someone, are more complicated than searching the web. For example, the assistants need to be able to understand the user’s intention so that they can select an appropriate action or give an appropriate answer. In addition, users expect the voice assistant to maintain the context of a dialog across the dialog. It has also been shown that users ask for more sophisticated information than they would look for on the web [4].

To improve search systems, an evaluation of the systems is necessary. For this evaluation, it is helpful to define a standard of correct answers [9]. With voice assistants, it is difficult to speak of correct answers because the answers are always personalized and contextual. For example, they are influenced by user location or previous user history [4]. Another way to evaluate search systems is to look at how long users stay on web pages or how often they click on a page [10]–[12]. However, it cannot be assumed that users will be dissatisfied if there is no interaction, because there are also voice assistants that offer users answers directly on the device so that they do not have to click on the device at all. In recent years, this “good” abandonment was considered in many studies [13]–[15]. As shown here, it’s not easy to measure users’ satisfaction with voice assistants.

B. Technology Acceptance Models

Davis’ established Technology Acceptance Model (TAM) has already laid the foundation for research into the acceptance of technologies. The model looks at how users accept and use a technology [16]. Initially used to assess the use of office software, the model has been applied to a multitude of different contexts.

The model is based on two prior models. The *Theory Of Reasoned Action (TRA)* [17] states that the individual intention determines whether a behaviour, i.e., the use of a technology, is carried out. Here, the attitude towards behaviour and subjective norms control the intention. The *Theory of Planned Behavior (TPB)* [18] also includes personal self-efficacy as an explanatory factor in the model.

According to the TAM, the actual application of a technology is decisively influenced on the one hand by the perceived usefulness and on the other hand by perceived ease of use. The attitude and perceived usefulness then determine the intention to use [16].

In addition to the TAM, the *Unified Theory of Acceptance and Use of Technology* (UTAUT, later UTAUT2) [19] is also an established model of technology acceptance. The model was published in 2012 by Venkatesh et al. as a further development of the TAM. According to it, four criteria mainly influence acceptance. These factors are *perceived usefulness*, *perceived ease of use*, *social influence* and *facilitating conditions* [19].

Both models consider individual criteria and their influence on the acceptance of a technology. In actual user decisions, however, the criteria are not considered separately, but the users weigh the various criteria against each other in order to decide whether they will use a technology or not. Therefore, these classical acceptance models cannot comprehensively map complex decision situations and a further methodological approach is necessary to relate the decision-relevant factors to each other [16], [19].

C. Acceptance of voice assistants

Voice assistants are today a hot, but also new topic and there has been little research about important factors for the adoption and use of voice assistants. Nevertheless, some interesting studies have already been published. Below we summarize the articles dealing with the acceptance of such assistants.

In their qualitative study, Kessler and Martin examined the acceptance of VAs for the first time. In their study, they extended the existing UTAUT2 model by the factors *data security*, *compatibility with other devices* and the *relationship to the device*. While *performance*, *hedonistic motivation* and *price* proved to be less significant, the newly introduced factors had a major impact on the acceptance of VAs [20].

Further, Arttu Kääriä considered the influence of *anthropomorphism*, i.e., the attribution of human characteristics, on the acceptance of such systems. He found that the *ease of use* and the *overall quality* of the system are much more important to users than *anthropomorphism*. For him one reason for this is that the quality is not the same for all systems, yet. It is not even sufficiently good at all, which is why factors such as *anthropomorphism* appear to be rather irrelevant [21]. This is likely to change in the future as VAs become better and better.

In contrast, Ahmadian et al. found that the *quality of interaction*, which includes anthropomorphism, has a positive effect on the acceptance. They combined some existent models for the adoption of new technologies and some social scientific concepts to a new model for the acceptance of voice assistants. In this model *information quality*, *system quality*, *interaction quality*, *trust* and *innovativeness* influence the acceptance of voice assistants positively. They found out, that the most important factor for the acceptance is the *interaction quality* and with it the anthropomorphism [22]. The two studies thus reach contradictory results in terms of whether anthropomorphism has an influence on the acceptance of voice assistants. However, only small samples were considered in both studies.

An online survey [23] showed that participants generally prefer to use a voice assistant in private locations. This holds particularly true for private information, that most users do not like to share aloud. The authors stated that the hesitancy to use voice assistants in public domains is possibly due to weak perceived social acceptability.

D. Privacy of voice assistants

The subject of privacy and security of voice assistants has been picked up early by researchers. Diao et al. showed that the built-in voice assistant (GVS) on Android phones might pose a threat to the users' privacy [24]: Malicious apps can play prepared audio files directly to the voice assistant and thus perform actions in the background without the users' knowledge. This way, permissions can be bypassed, leaving sensitive information vulnerable to attacks.

Alepsis and Patsakis found that Android devices that are not locked by PIN or pattern are particularly vulnerable to data abuse. For example, users can download apps from the Appstore that activate a voice assistant without telling the user. The installed assistants can also be used to share personal data. This happens for example via audio recordings.

Alepsis and Patsakis also point out that in the age of the Internet of Things, a compromised phone can be used to attack other nearby devices with voice control and that Smartwatches could be used to unlock devices and thus increase the possibility of accessing users' data. Even training voice assistants on certain voices does not always protect them from such attacks, because they can also be bypassed by an adapted artificial language module. Accordingly, this vulnerability is more dangerous than was initially assumed [25].

When contrasting these privacy and security focusing studies with the prior acceptance studies, it becomes apparent that the focus of the study might sway the answers of its participant in the direction implicitly stated in the study. When asked directly whether users would like to share intimate photos of themselves with a large corporation, many would wholeheartedly disagree, still millions do so everyday on social media. The underlying privacy paradox is at play, counteracting thoughtful rational choices with impulsive affect-based actions. Some studies therefore focused their aim at the privacy-utility trade-off [26]–[30]. These studies still ask the rational mind of the survey participant, yet, taking this trade-off into account for the use of some example connected technology, yielded more realistic results.

In our study, we aim to address this trade-off as well, trying to understand how (possibly) contradicting factors such as price, quality, and privacy of a voice assistant play together in forming acceptance of these devices.

III. Method

Since there has been little research on the adoption of voice assistants so far, we have chosen an approach for our study that enables to balance factors that can determine the acceptance of voice assistants against each other. For this, we conducted a conjoint analysis and combined it with a few additional questions in a survey. Conjoint studies are useful when trade-offs in the mind of participants should be measured.

A. Conjoint Analysis

(CA) In order to understand how we designed the survey, we first must shortly digress and explain how conjoint measurement and analysis work. CA was developed in the 1960s by Luce and Tukey and is a quantitative empirical research method that combines a measurement model with a statistical estimation algorithm. The method aims to study consumer choices or preferences for complex products. For this, the influence of individual product features is considered [6]. Conjoint analyses thus allow to measure multiple attributes (e.g., color of a product and price) simultaneously, what classical survey methods do not allow. Participants or potential users can choose the product alternative that is the best in their eyes. Individual attributes are then differentiated in levels (e.g., red, blue, yellow or 10, 20, and 30 Euros). Product properties are thus ordered and weighted and it also becomes visible which compromises between the individual product properties are acceptable to the users. In addition, the participants can be segmented into user groups. For factors with interacting influences conjoint analyses yield better models.

B. Choice-based Conjoint Analysis

In this study, we used the Choice-Based-Conjoint Analysis (CBC)[7], [8], [31], which is the most widely used conjoint related technique. Previous types of conjoint analysis required participants to rate or rank products or concepts, and this was used to determine their preference. In contrast, in CBC-analysis, the participant selects one of the proposed products.

We used the CBC-analysis to examine how much the individual attributes or attribute levels affect the acceptance or rejection of VAs [7], [8], [31]. Here, the participants were shown one after another different voice assistants consisting of three attributes in different levels [31].

In a CBC-analysis, the participants repeatedly choose one of several products, which consists of different attributes and different levels. Thus, a model for the selection probabilities is calculated using multinominal logit or probit models [31]. In our study, users had to choose a preferred VA in 13 choice tasks [31]. In each task, we displayed three VAs and the participants should choose the one they like best. Which VAs the participants see, or how the VAs are configured, was randomly chosen. With a full-factorial design, all participants see all possible combinations. In our study a full-factorial design would require 27 decision tasks ($3 \times 3 \times 3$). To ensure that the survey is not too time-consuming for the participants, we reduced the number of decision tasks to 13 per participant. This means that every single user has not rated every possible combination and that usually several users have not seen the same choice set. Nonetheless, the results are 99% comparable with the results of an orthogonal design. And the standard error of the design, which tells how accurate the main effects are, was below the 0.05 limit. The smaller it is , the better. Due to previous studies, it is recommended that the standard error for each attribute layer should be less than 0.05 [32].

In CBC-analysis, the relative importance of the attributes and the part-worth utilities of the attribute levels are usually calculated. To calculate the part-worth utilities for each attribute level separately, we used Hierarchical-Bayes (HB)-estimation. In this estimation, the personal part-worth utilities are combined with the average of the overall sample to get part-worth utilities, which indicate the attractiveness of the levels in comparison to the other levels within the same attribute [33]. The advantage of this is that the results or the computed utilities are also reliable for small samples or when the participants answer only a set of the possible decision sets.

The importance of the individual attributes is also determined using the part-worth utilities. The higher the importances of an attribute the higher is the effect for the decision task—in our case the selection of a VA. In order to calculate the relative importances, the range of the part-worth utilities of one attribute must be normalized by the total range of the part-worth utilities of all attributes. The relative importance show how much the attributes influence the decision of the participants (e.g. which attribute has the highest influence) and if they have a positive or negative influence on the decision [7], [34]. The described preferences show the acceptance of a technology such as virtual voice assistants in our example [8].

C. Identification of acceptance relevant factors

Subsequent to a literature review, we identified three relevant factors for the acceptance of VAs: *NLP-performance*, *price*, and *privacy*. Here, we explain in detail why we integrated exactly these attributes in the conjoint analysis.

Today, several hundred million cameras are installed in private and public environments worldwide [35], [36]. Among other purposes, these cameras serve to increase the security of people in their everyday lives. However, critics warn that the use of cameras, microphones, and localization technologies—thus recording and storing data—is an invasion of privacy [37]. So there is a dichotomy between the advantages and disadvantages of such technologies, which can also be important for VAs. Therefore, it is important to investigate if and when users will feel hurt when using VAs in their private sphere. As shown in section II, privacy has proven to be an important criterion for the acceptance of VAs. In this study, we are interested in how individuals weigh privacy against other aspects. For example, we are particularly interested in the trade-off between privacy and price, which has already been considered in some studies, but not for VAs. As this trade-off has often shown, price always plays an important role in the acceptance of new technical products. It is especially interesting to us, if individuals trade their privacy for cheaper products or, on the contrary, if they are willing to pay more for a VA in order to protect their privacy. Aside from this well-known trade-off, we wanted to investigate two more trade-offs:

We assumed a possible trade-off between *price* and *NLP-performance* as it is plausible that individuals are willing to pay more money for a higher functionality. On the other hand we assume that people are afraid of virtual voice assistants that can understand them totally, because of the already showed importance of privacy concerns.

Figure 1 shows the research design of this study. As it shows we looked at some user factors and the above described attributes that contain each three levels. Further we investigated the influence of the attributes on the self-reported acceptance of voice assistants in the private and professional environment.

D. Attributes and levels

In this study, we used a conjoint analysis with three attributes, each with three levels, that are introduced here. *Natural language processing performance (NLP-performance)*: The first attribute we considered is *NLP-performance*. For this attribute, we contrasted three different levels of NLP-performance, each of which represents a further technical development or an improvement in functionality. The first level *simple commands* (level 1) refers to a voice assistant that only recognizes fixed instruction sets and can execute simple commands. This level corresponds to the quality of voice assistants from about five years ago. The next level or enhancement *advanced instructions* (level 2) means that the VA understands a range of commands and may work context-based in some circumstances. However, the VA can not work context-based in every situation. This is how the quality of virtual language assistants is today. The next level *natural language* (level 3) refers to a VA allowing natural language and executing even awkwardly formulated instructions. At this level the VA fully understands the user as another human being would. This quality is expected in a few years in the future.

Price The second attribute we have chosen is the *price* of a voice assistant. Here, we contrasted again three levels and we have subdivided the *price* into *50 Euro* (level 1), *100 Euro* (level 2) and *200 Euro* (level 3). These levels were chosen as these represent typical prices of home VA systems. We are aware that these prizes are highly subsidized products, but they accurately depict the current market situation.

Privacy Additionally, we integrated three different levels of *privacy* in our study: *offline* (level 1), *keyword* (level 2) and *always on* (level 3). The third level means that the VA always transmits what he records to the manufacturer for evaluation. At level 2 the VA transmits the recorded data of a conversation only after recognizing a keyword. Level 1 says that the VA analyzes the recorded data locally and only transmits data that is necessary for the execution of the commands.

E. Questionnaire Design

Our survey consisted of three parts: The first part dealt with user-related factors. Here, we queried as demographics *age* and *gender*. Further, we asked whether the participants have *previous experience in dealing with VAs* on a six-point Likert scale from 1 = none to 6 = very much. Last, we surveyed the participants' *self-efficacy in interacting with technologies*. In the second part we asked the participants about their acceptance of VAs in their private and professional environment and in the third part we looked at three acceptance-relevant factors of VAs using a conjoint analysis.

Self-Efficacy in Interacting with Technology (SET): New technologies will not be adopted until potential users know

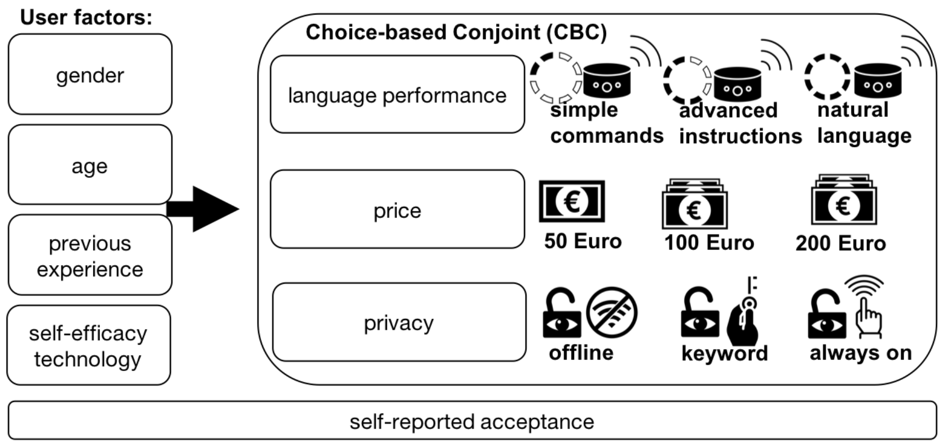

Fig. 1. Model of the conducted user study. User factors are established before a set of 13 conjoint tasks using three attributes. Self-reported acceptance is measured last.

how to use the technology properly [38]. VAs also raise the question of whether people understand how to deal with them and whether people accept the technology. SET measures whether people are generally open to new technologies and whether they can interact successfully with technologies. SET can refer to both medical [39] and non-medical [40] contexts. We measured the participant's self-efficacy in interacting with technology on a 5-item scale, which can be seen in Table I and the scale achieved a good internal reliability ($\alpha = .876$).

TABLE I
ITEMS OF THE SCALE SELF-EFFICACY IN INTERACTING WITH TECHNOLOGY (SET).

	Self-Efficacy in Interacting with Technology: Items (n=93; $\alpha = .819$)
1	Dealing with technology is a pleasure.
2	I don't trust technique in general.
3	I find it easy to deal with technology.
4	I am interested in technology.
5	I avoid technology, if it is possible.

Self-reported acceptance of virtual voice assistants To find out whether users would accept or reject virtual voice assistants in their private and professional environments, we asked how the participants would react between welcoming and protesting the decision that someone installs a VA using a graphical slider. We first asked how they would react if a close relative wanted to permanently install a voice assistant in their shared living room. Second, we asked how they would react if a VA were to be permanently installed at work.

F. Data Acquisition and Analysis

Our web-based survey was distributed via mailing lists, instant messaging services, and personal contacts during June and July 2018 and February 2019. We used parametric and non-parametric methods to analyze our results. We calculated bivariate correlations using Pearson's r or Spearman's ρ. Further, we computed univariate analyses of variance (ANOVA). We set the level of significance to $\alpha = .05$. We analyzed the results in the Sawtooth Software using Hierarchical-Bayes (HB) estimation and first calculated the relative importances and then the part-worth utilities of the attributes. For arithmetic means (M) we report the 95%-confidence intervals; denoted as $[lower, upper]$.

IV. RESULTS

In order to contextualize our results, we first present a short sample description. This is followed by the relative importances and the part-worth utilities of the three attributes (see section III-A) for the overall sample. Lastly, we describe the relative importances of the attributes for the four user groups, which we identified using clustering methods.

From our 93 respondents 53% (49) were female and 47% (44) were male. On average the participants were 28.9 years old with a standard deviation of 10.5 years. The participants had little previous experience with voice assistants (VAs) ($M = 2.52, SD = 1.21$, 1=no previous experience, 6=very much previous experience). Men tend to have more *previous experience* ($M = 2.80, SD = 1.46$) in dealing with virtual voice assistants than women ($M = 2.27, SD = 0.88$). We also found an effect of *age* on the *previous experience* ($r = -.347, p = .001$). Younger people have more *previous experience* than older people.

Overall, our participants are rather technically affine ($M = 3.20, SD = 0.61$, 1=does not apply, 4=does apply). Men ($n = 49, M = 3.36, SD = 0.54$) reported a slightly larger technology affinity ($r = -.248, p = .017$) than women ($n = 49, M = 3.06, SD = 0.65$).

When asked how the participants would react if someone were to install a virtual voice assistant without asking, most participants would protest. Here, older people are more likely to protest against the installation of a voice assistant. This applies to both the private ($r = .289, p = .005$) and the professional ($r = .346, p = .001$) environment. Further, we found a significant negative effect of *previous experience* on the acceptance of voice assistants in the private ($r = -.563, p < .001$)

and professional environment ($r = -.301, p = .003$). The more experience users have with voice assistant the less likely they are to accept them.

A. Attractiveness of a virtual voice assistant

Regarding the relative importances ($r.i.$, see Figure 2) of the conjoint analysis we found out that *language performance* ($r.i. = 19.91$) and *price* ($r.i. = 20.50$) are almost equally important to our participants, while *privacy* ($r.i. = 59.59$) is much more important.

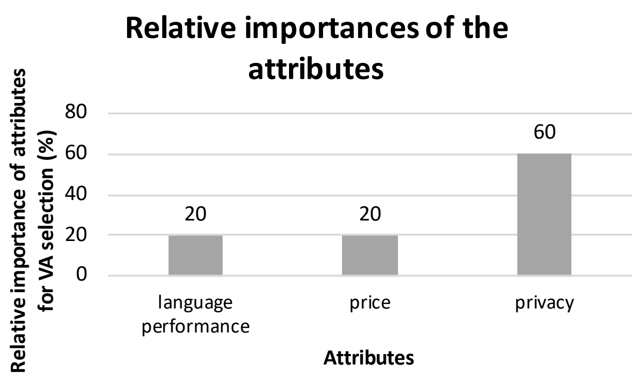

Fig. 2. Relative importance of the three attributes. The value indicates how important the attribute is in joint decision making.

The part-worth utilities (see Figure 3) also showed that *privacy* is the most important factor, because the differences between *always on* and *offline* ($\Delta = 150.50$) and between *always on* and *key word* ($\Delta = 135.28$) are clearly larger than the differences between the levels of the other attributes. The participants thus do not want the assistant to always be on, but it is less important to them, if the assistant is offline or reacts to a keyword. The differences within the other attributes are smaller. There is seemingly one undesirable option in terms of *price*, because the difference between *200 Euro* and *100 Euro* ($\Delta = 40.54$) and *200 Euro* and *50 Euro* ($\Delta = 51.67$) are clearly bigger than the difference between *100 Euro* and *50 Euro* ($\Delta = 11.13$). For *language performance*, the distances between the levels are similarly large, which means that the participants prefer *natural language* in comparison to *advanced instructions* ($\Delta = 21.74$) similarly strong as they prefer *advanced instructions* in comparison to *simple commands* ($\Delta = 28.49$).

B. User Segments of virtual voice assistants

After the CBC-analysis of the overall sample we took a closer look at the participants and identified possible user groups. We computed a latent-class-analysis [41] and used the three attributes of the voice assistant *NLP-performance*, *price*, and *privacy* to divide the participants into user segments of possible target groups of voice assistants.

Using *latent-class-analysis* we found four different clustered solutions. To check the stability of the solutions, we calculated five solution iterations from different random starting points

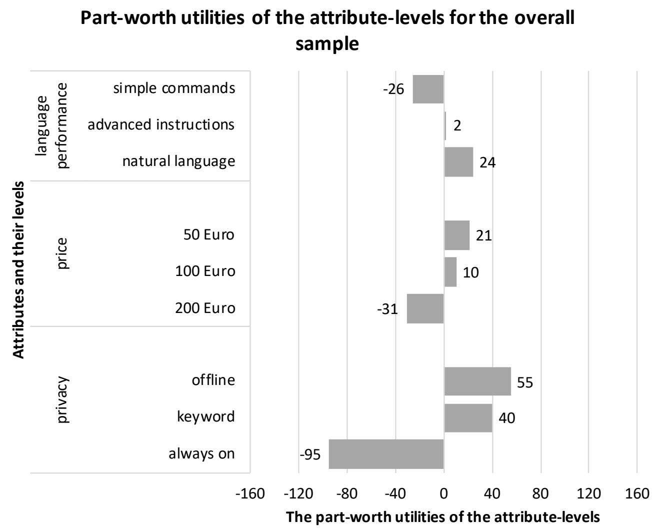

Fig. 3. Part-worth utilities of all levels of the three attributes

as input values and compared the results [41]. Our decision for the 4-group solution—among other reasons—was based on the *Consistent Akaike Information Criterion (CAIC)*, for which applies the smaller this value is, the better [42]. Compared to the other solutions (see Table II), the 4-group solution with CAIC = 1547.0 received the lowest values and thus showed the best data fit. Patterns in the values can also be regarded as a further measure for the selection of the number of groups. The other values of the model quality (*Chi-Square*, *Bayesian Information Criterion (BIC)*, *Adjusted Bayesian Info Criterion (ABIC)* see Table II) showed larger differences between the 2- and 3-group solution and the 3- and 4-group solution, but hardly changed afterwards for the 5-group solution. Such patterns are indications for a good choice of the number of groups [41]. Each attribute turned out to be significant and all respondents can be allocated to one of the groups with 96.77% accuracy. In addition to the statistical quality we looked at the best option regarding the considered content for our decision for the 4-group solution. For three groups *privacy* is especially important, which reflects the overall sample. Still there are two groups (*1* and *3*) for which privacy is almost the only important thing, whereas one group (*4*) who consider the other two attributes to be at least a little more important as well. *group 2* is particularly interesting against the background of the considered trade-offs, because as they appear to be willing to release their data for a lower price. Also *group 4* is interesting due to the trade-off between *NLP-performance* and *privacy*.

With the four groups we can see which factors are motivating and inhibiting for each group so that various possible target groups of voice assistants emerge. Following, we describe each group individually and the respective preferred voice assistants or the motivating and inhibiting factors.

1) Group 1: data protectors: As Figure 4 shows, the *data protectors* (12% of all participants) attach almost all

TABLE II
DATA FIT CRITERIA OF DIFFERENT SOLUTIONS.

Solution	CAIC	Chi-Square	BIC	ABIC
2 groups	1638.7	1120.8*	1625.7	1584.4
3 groups	1587.4	1228.8	1567.4	1503.8
4 groups	1547.0*	1325.9	1520.0*	1434.2
5 groups	1564.2	1365.4	1530.2	1422.2*

Asterisks* indicate the best model fit by criterion.

importance to *privacy*. The relative importance of *privacy* is with 86% clearly higher than the relative importances of the other attributes, which only reach up to 12%. The relative importance of *privacy* is also higher for the *data protectors* than for the other groups and clearly higher than for the whole sample (see Figure 2; $r.i. = 60$). Further the *data protectors* are the only group for which the *price* is irrelevant, at least compared to the other attributes.

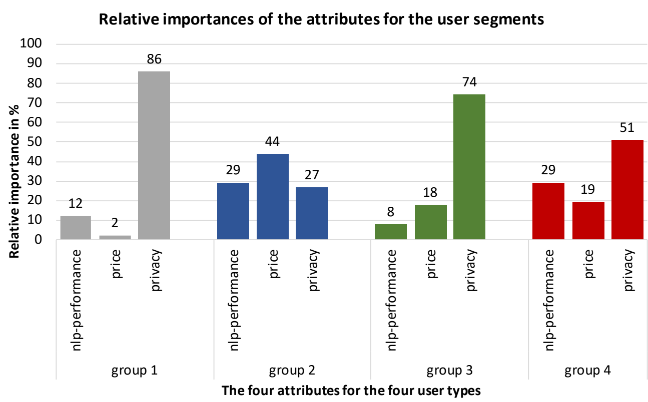

Fig. 4. Relative importances of the three attributes for the four groups. Grey = data protectors, blue = thrifts, green = thrifty data protectors, red = limited data protectors.

We computed a Welch's t-test for unequal variances (Levene's test: $F(3, 89) = 8.48$, $p < .001$) and integrated the *clusters* or *groups* as independent and the relative importance of *privacy* as dependent variable. We found that at least two groups of the segments differ significantly with a large effect ($F(3, 30.25) = 62.57$, $p < .001$; $\eta^2 = .72$). The Games-Howell post-hoc test showed that *group 1* rates *privacy* significantly more important than *group 2* (33.18, CI$[22.12, 44.23]$[1], $p < .001$, $cp < .001$) and *group 4*. (15.25, CI$[6.72, 23.78]$, $p < .001$, $cp < .001$), but it did not show a significant difference to *group 3* ($p > .05$). *Price* is significantly less important for *group 1* than for the other groups.

2) Group 2: thrifts: For the *thrifts* (18%) the *price* is the most important (see Figure 4). The distance to the second most important factor is significantly smaller compared to *group 1* and *group 3*. They also rate the other two attributes as relatively important and almost equal important. Together with *group 4* and in contrast to the other groups, they rate *NLP-performance* as relatively important. For both groups, the attribute achieves a relative importance of 29%. Further the *thrifts* are the only group, who do not rate *privacy* as most important, but *price*. They rate the *price* also as clearly more important than the overall sample (see Figure 2; $r.i. = 21$) and still clearly more important than the other attributes.

For *group 2* the Games-Howell post-hoc test after the Welch's t-test showed that *privacy* is significantly less important for them than for the other groups (*group 1*: s.a.) ; group 3: -32.46, CI[-41.17, -23.75], $p < .001$, $cp < .001$), group 4: -17.92, CI[-26.67, -9.18], $p < .001$, $cp < .001$). Further we computed a Welch's t-test (Levene's test: $F(3, 89) = 13.96$, $p < .001$) with the clusters as independent and *NLP-performance* as dependent variable. It showed that at least two groups differ significantly ($F(3, 29.79) = 35.97$, $p < .001$, $\eta^2 = .68$). According to the Games-Howell post-hoc test, *NLP-performance* is significantly more important for *group 2* than for *group 3* (15.40, CI[5.80, 25.00], $p = .001$). For *group 2 price* is significantly more important than for all the other groups.

3) Group 3: thrifty data protectors: Similarly to the *data protectors* the *thrifty data protectors* (33%) clearly rate *privacy* as the most important factor (see Figure 4; $r.i. = 74$) and as well clearly more important than the overall sample ($r.i. = 60$). In contrast to *group 1* they take the *price* of a virtual voice assistant a bit more into their evaluation.

Likewise, for *group 1*, the post-hoc test showed that *privacy* is significantly more important for *group 3* than for *group 2* (see above) and *group 4* (14.54, CI[10.96, 18.12], $p < .001$, $cp < .001$). For *group 3 NLP-performance* is significantly less important than for *group 2* and *group 4*. For *group 3 price* is significantly more important than for *group 1* and significantly less important than for *group 2*.

4) Group 4: limited data protectors: Although *privacy* is most important for *the limited data protectors* (37%), as is the case for *group 2*, the distance to the other attributes is less clear than for *group 1* and *group 3*. In addition to *privacy*, *NLP-performance* and *price* are also important for *group 4*, whereby they rate the *NLP-performance* as slightly more important than the *price*.

According to the post-hoc test *group 4* rates *privacy* significantly more important than *group 2*, but significantly less important than *group 1* and *group 3*. Further for *group 4* is *NLP-performance* significantly more important than for group 3 (13.97, CI[10.31, 17.64], $p < .001$). *Group 4* rates the *price* as significantly more important than *group 1* and significantly less important than *group 2*.

C. User Segments, personality and acceptance of VAs

The four segments differ in personality traits and acceptance of voice assistants (see Figure 5). To find out if the groups differ in their characteristics and in how motivated they are to use a voice assistant, we conducted a MANOVA with the clusters as independent and the user characteristics and acceptance-variables as dependent variables. We found an overall significant difference between the groups ($F(18, 258) = 2.47$, $p = .001$; Wilk's $\lambda = 0.611$, partial $\eta^2 = .15$). Looking at the

[1]Numbers in square brackets refer to to 95% confidence interval

individual characteristics and acceptance factors, the groups differ significantly in their previous experience with voice assistants ($F(3, 89) = 3.66$, $p = .015$; partial $\eta^2 = .11$), their private acceptance ($F(3, 89) = 8.89$, $p < .001$; partial $\eta^2 = .23$) and professional acceptance ($F(3, 89) = 5.43$, $p = .002$; partial $\eta^2 = .15$). The Tukey-HSD post-hoc test showed that *group 4* has significantly more previous experience with voice assistants (.94, CI[.18, 1.70], $p = .009$). The Tukey-HSD post-hoc test for the *private acceptance* showed that *group 1* would significantly stronger accept a voice assistant in their private environment than *group 4* (2.04, CI[.02, 4.07], $p = .046$) and *group 3* showed a higher acceptance in comparison to *group 2* (1.88, CI[.11, 3.66], $p = .033$) and *group 4* (2.77, CI[1.32, 4.23], $p < .001$). Regarding the professional environment, *group 2* would accept a voice assistant significantly stronger than *group 4* (1.87, CI[.23, 3.50], $p = .019$) and *group 3* would accept a voice assistant in their professional environment stronger than *group 4* (1.81, CI[.44, 3.19], $p = .005$).

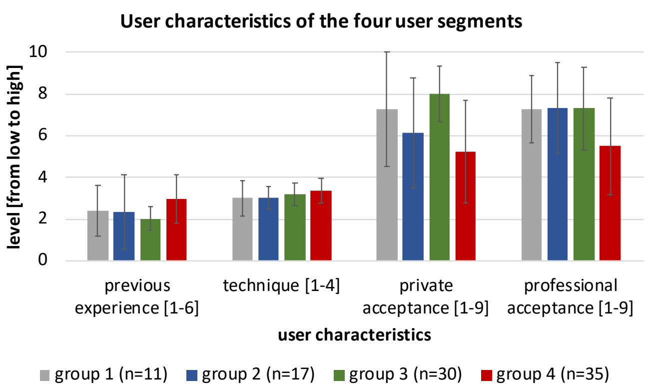

Fig. 5. User characteristics and acceptance factors for the four user segments. Error bars denote 95% confidence intervals.

V. Discussion, Limitations and Future Work

The results show that not the price of a voice assistant (VA) is the most important factor for its acceptance, but privacy. Most participants definitely do not want a voice assistant that is always on and always transmitting data. The decisions of our participants relate well with a number of studies looking at the individual privacy-utility trade-off in the use of technology [26]–[30].

However, the highest price as well as simple commands and advanced instructions are also rather rejected. But, is there a perfect voice assistant or a best compromise for (potential) users? The evaluation of the price is usually dependent on a cost-benefit trade-off, but it is conceivable that the benefit to the participants was not clear enough in the privacy settings and voice features. In the future, we will review the results using a larger sample and possibly other factors.

Contrary to expectations the desire for a better NLP-perfomance does not mean that the (most) participants would also pay more money for a voice assistant.

Within this conjoint study, we looked at several trade-offs. When asking the participants about the price in relation to the NLP-performance and privacy, these trade-offs can be affected by different socio-economic statuses. It is conceivable, that better-off workers are rather able and thus willing to pay more money for a voice assistant than people with less money. Hence, the price could be less important for some people than for others. This could have biased our results and has to be taken into account for the further consideration of our results. Nevertheless, looking at trade-offs with the price, in our sample, the price seems to be less important than privacy.

Within the attribute NLP-performane we expected that participants would reject “full” natural language. We assumed this, because we thought that individuals are afraid that such systems can understand them all-embracing. Also, against the background that the participants do not want the assistant to always be online, it would be expected that they would rather reject this option. In our study, the participants apparently actually perceived natural language as a better quality and did not rate it negatively. It is also conceivable, however, that the test persons rated the natural language so positively, since we have directly asked about how it works and a better function is therefore automatically perceived more positively. In the future, it would therefore be interesting to implicitly interrogate the perception of natural language or to test it in an wizard-of-oz style experiment.

Further, there is no perfect voice assistant, but different potential user groups with different preferences. Three of four groups rated the price as clearly less important than privacy. Still, the price is important for group 2.

In future we would like to consider additional factors that could have an influence on the acceptance of virtual voice assistants. Further, we plan to conduct a study with a larger sample, in order to verify the here identified target groups.

VI. Conclusion

In this study, we exploratorily investigated the effect of natural language processing performance, price, and privacy on the acceptance of virtual voice assistants. Privacy turned out as the most important aspect for the acceptance of voice assistants. But, there are also unwanted options for price and NLP-performance and there is a possibly smaller group of the population that desires particularly a low price. Our results indicate that for different individuals different aspects lead to the acceptance of virtual voice assistants. We identified four different user segments and we could not find a perfect combination of the investigated factors for all potential users, but we have found different voice assistants that appeal to different potential user groups. Future research, using similar methods and experimental validation, will be help understand both human privacy behavior in homes and offices, as well as help design virtual assistants, whose omnipresence seems inevitable in the future.

Acknowledgment

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

REFERENCES

[1] M. B. Hoy, "Alexa, siri, cortana, and more: An introduction to voice assistants," *Medical reference services quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[2] A. C.-S. Bogazici and S. M. Bogazici, "Understanding the Adoption of Voice Activated Personal Assistants," *International Journal of E-Services and Mobile Applications*, vol. 3, no. 9, pp. 1–21, 2017.

[3] Google Inc. (2015). Teens use voice search most, even in bathroom, google's mobile voice study finds, 2015, [Online]. Available: http://prn.to/1sfiQRr.

[4] J. Kiseleva, K. Williams, A. H. Awadallah, A. C. Crook, I. Zitouni, and T. Anastasakos, "Understanding User Satisfaction with Intelligent Assistants," *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval - CHIIR '16*, pp. 121–130, 2016.

[5] M. Negri, M. Turchi, J. G. C. de Souza, and D. Falavigna, "Quality estimation for automatic speech recognition," *COLING*, pp. 1813–1823, 2014.

[6] R. Luce and J. W. Tukey, "Simultaneous conjoint measurement: A new type of fundamental measurement," *Journal of Mathematical Psychology*, no. 1, pp. 1–27, 1964.

[7] B. Orme, "Interpreting the results of conjoint analysis, getting started with conjoint analysis: Strategies for product design and pricing research," *Research Publications LLC*, pp. 77–89, 2010.

[8] K. Arning, "Conjoint Measurement," in *International Encyclopedia of Communication Research Methods.* J. M. Wiley-Blackwell, Ed., JohnWiley and Sons, 2017.

[9] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," *ACM Trans. Inf. Syst. (TOIS)*, vol. 4, no. 20, pp. 422–446, 2002.

[10] E. Agichtein, E. Brill, and S. T. Dumais, "Improving web search ranking by incorporating user behavior information," *SIGIR*, pp. 19–26, 2006.

[11] S.Fox, K.Karnawat, M.Mydland, S.T.Dumais, and T.White, "Evaluating implicit measures to improve web search," *ACM Trans. Inf. Syst. (TOIS)*, vol. 2, no. 23, pp. 147–168, 2005.

[12] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay., "Accurately interpreting clickthrough data as implicit feedback," *SIGIR*, pp. 154–161, 2005.

[13] A. Chuklin and P. Serdyukov, "How query extensions reflect search result abandonments," *SIGIR*, pp. 1087–1088, 2012.

[14] ——, "Good abandonments in factoid queries," *WWW (Companion Volume)*, pp. 483–484, 2012.

[15] J. Li, S. B. Huffman, and A. Tokuda, "Good abandonment in mobile and pc internet search," *SIGIR*, pp. 43–50, 2009.

[16] F. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, no. 13, pp. 319–340, 1989.

[17] M. Fishbein and I. Ajzen, "Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research," *Addison-Wesley Publishing Company, Inc.: Reading*, 1975.

[18] F. Davis, "The Theory of Planned Behavior," *Organ. Behav. Hum. Decis. Process.*, no. 50, pp. 179–211, 1991.

[19] V. Venkatesh, J. Thong, and X. Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Q.*, no. 36, pp. 157–178, 2012.

[20] S. K. K. &. M. Martin, "How do potential users perceive the adoption of new technologies within the field of Artificial Intelligence and Internet-of-Things? A revision of the UTAUT 2 model using Voice Assistants," 2017.

[21] A. Kääriä, "Technology acceptance of voice assistants: anthropomorphism as factor," 2017.

[22] M. Ahmadian, O.-K. D. Lee, and F. Nasirian, "Ai-based voice assistant systems: Evaluating from the interaction and trust perspectives," *Twenty-third Americas Conference on Information Systems, Boston, 2017*, 2017.

[23] A. Easwara Moorthy and K.-P. L. Vu, "Voice Activated Personal Assistant: Acceptability of Use in the Public Space," in *HIMI 2014: Human Interface and the Management of Information. Information and Knowledge in Applications and Services*, Springer, Cham, 2014, pp. 324–334. DOI: 10.1007/978-3-319-07863-2_32. [Online]. Available: http://link.springer.com/10.1007/978-3-319-07863-2%7B%5C_%7D32.

[24] W. Diao, X. Liu, Z. Zhou, and K. Zhang, "Your Voice Assistant is Mine," in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices - SPSM '14*, New York, New York, USA: ACM Press, 2014, pp. 63–74, ISBN: 9781450331555. DOI: 10.1145/2666620.2666623. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2666620.2666623.

[25] E. Alepis and C. Patsakis, "Monkey says, monkey does: Security and privacy on voice assistants," *IEEE Access*, vol. 5, pp. 17 841–17 851, 2017.

[26] A. C. Valdez and M. Ziefle, "The users' perspective on the privacy-utility trade-offs in health recommender systems," *International Journal of Human-Computer Studies*, 2018.

[27] R. K. Chellappa and R. G. Sin, "Personalization versus Privacy: An Empirical Examination of the Online Consumers Dilemma," *Information Technology and Management*, no. 1, pp. 181–202, 2005.

[28] K. Mabley, "Privacy vs. Personalization: Part III.," *Cyber Dialogue, Inc.*, 2000.

[29] D. Taylor, D. Davis, and R. Jillapalli, "Privacy concern and online personalization: The moderating effects of in- formation control and compensation," *Electronic Commerce Res.*, no. 3, pp. 203–223, 2009.

[30] L. Burbach, J. Nakayama, N. Plettenberg, M. Ziefle, and A. Calero Valdez, "User preferences in recommendation algorithms: The influence of user diversity, trust, and product category on privacy perceptions in recommender algorithms," in *Proceedings of the 12th ACM Conference on Recommender Systems*, ACM, 2018, pp. 306–310.

[31] V. R. Rao, *Applied Conjoint Analysis*. Springer, 2014.

[32] S. Software, *Cbc tutorial and example*, 2019. [Online]. Available: https://www.sawtoothsoftware.com/help/lighthouse-studio/manual/index.html?cbctutorialandexample.html.

[33] B. Orme, "Sawtooth Software MaxDiff Analysis : Simple Counting, Individual-Level Logit, and HB," *Sawtooth Software, Inc. Research paper series*, vol. 98382, no. 360, 2009.

[34] K. Chrzan and B. K. Orme, "An overview and comparison of design strategies for choice-based conjoint analysis," presented at the Proceedings of the Sawtooth Software Conference, Sawtooth Software Research Paper Series, 2000, pp. 161–177.

[35] K. Dailey. (2013). The rise of cctv surveillance in the us, [Online]. Available: http://www.bbc.com/news/magazine-22274770.

[36] N. La Vigne and S. Lowry, "Evaluation of camera use to prevent crime in commuter parking facilities: a randomized controlled trial," *Technical report. Urban Institute, Washington, DC*, 2011.

[37] R. Whitaker, "The End of Privacy: How Total Surveillance is Becoming a Reality," *The New Press, New York*, 1999.

[38] N. &. B. Z. Krömer, "Von eHealth zu mHealth: Bedingungen und Barrieren für eine Nutzung mobiler Gesundheitsangebote," *Gesundheitskommunikation als-Forschungsfeld der Kommunikations- und Medienwissenschaft*, pp. 193–207, 2014.

[39] Y. Gao, H. Li, and Y. Luo, "An Empirical Study of Wearable Technology Acceptance in Healthcare," *Industrial Management & Data Systems*, vol. 115, no. 9, pp. 1704–1723, 2015. DOI: 10.1108/IMDS-03-2015-0087.

[40] K. Arning and M. Ziefle, "Understanding Age Differences in PDA Acceptance and Performance," *Computers in Human Behavior*, vol. 23, no. 6, pp. 2904–2927, 2007.

[41] 2. Sawtooth Software, "Survey Software & Conjoint Analysis - CBC Latent Class," *Technical Paper. Technical Paper Series. Retrieved at 2nd March from https://www.sawtoothsoftware.com/support/technical-papers/sawtooth-software-products/cbc-latent-class-technical-paper-2004*, 2004.

[42] T. Teichert, *Nutzenschätzung in Conjoint-Analysen. Theoretische Fundierung und empirische Aussagekraft. First edition*. Tremblay, 2001.

About the Authors

Laura Burbach (M.A.) is research assistant at the Human-Computer Interaction Center. Since 2018 she is part of the junior research group 'Digitale Mündigkeit'. She is currently investigating whether and to what extent different recommendation systems, voice assistants and individuals are accepted by users of social media. Besides her research focuses on target groups of Life-Logging.

Patrick Halbach (M. Sc.) works as research assistant at the chair of Communication Science at RWTH Aachen University, Germany and is part of the junior research group 'Digitale Mündigkeit'. In his research, he currently deals with opinion formation in online contexts, especially focusing on the impact of social media and the spread of misinformation on such platforms.

Nils Plettenberg (B. Sc.) works as student assistant at the chair for communication science at the RWTH Aachen University, Germany where he is part of the junior research group 'Digitale Mündigkeit'. He is studying the master's program electrical engineering with a major in computer engineering. His research focuses on the spread of news in social media.

Johannes Nakayama (B. Sc.) is a student assistant at the chair for cimmunication science at RWTH Aachen University and a part of the junior research group "Digitale Mündigkeit". He is a Master's student majoring in technical communication (communication science and computer science) and he specializes in recommender systems research.

Martina Ziefle is Professor of Communication Science and head of the Human Computer Interaction Center of RWTH Aachen University. Her research focuses on the interface between humans and technology, taking into account different usage contexts and user requirements. She focuses in particular on the sensitive area of eHealth technologies, in which user diversity and technology acceptance play a decisive role.

André Calero Valdez (PhD) is junior research group leader at the RWTH Aachen University, Germany. His research focuses the interaction of humans and algorithms such as recommender systems, information visualization, and machine learning in a wide variety of applications. The fields of application range from social media over health informatics to industry 4.0.