Design Space and Users' Preferences for Smartglasses Graphical Menus: A Vignette Study

Radu-Daniel Vatavu
MintViz Lab | MANSiD Research Center
University Ştefan cel Mare of Suceava
Suceava, Romania
radu.vatavu@usm.ro

Jean Vanderdonckt
LouRIM
Université catholique de Louvain (UCLouvain)
Louvain-la-Neuve, Belgium
jean.vanderdonckt@uclouvain.be

ABSTRACT

We address in this work visual menus for smartglasses by proposing a wide range of design options, such as shape, location, orientation, and presentation modalities, which we compile in a design space with eight dimensions. From this design space, we select a subset of fourteen 2-D menus, for which we collect users' preferences during a vignette experiment with N=251 participants. We report numerical measures of absolute, relative, and aggregate preference for smartglasses menus, and employ a particular Thurstone's pairwise comparison technique, the Bradley-Terry model, to evaluate menu designs. Our results highlight key variables influencing users' preferences regarding the visual appearance of smartglasses menus, which we use to discuss opportunities for future work.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile devices**; **Graphical user interfaces**; **Displays and imagers**.

KEYWORDS

smartglasses, user study, A/B testing, vignette study, 2-D menus, visual aesthetics, preference measures, Bradley-Terry model

ACM Reference Format:
Radu-Daniel Vatavu and Jean Vanderdonckt. 2020. Design Space and Users' Preferences for Smartglasses Graphical Menus: A Vignette Study. In *19th International Conference on Mobile and Ubiquitous Multimedia (MUM 2020), November 22–25, 2020, Essen, Germany*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3428361.3428467

1 INTRODUCTION

Advances in wearable computing and Augmented/Mixed Reality (AR/MR) technology have rendered smartglasses devices available to mass consumers with forecasts projecting 22 million units to be shipped worldwide by the year 2022 [67] that will generate a total expected revenue of 19.7 billion US dollars [66]. Smartglasses designs range from Head-Mounted Displays (HMD) that render photorealistic computer-generated graphics in the form of holographic computing [47,52] to fashionable tech eyewear, such as "Spectacles by Snapchat" [64], and to miniaturized gadgets with micro video cameras embedded into their temples, impossible to distinguish from regular eyeglasses by the uninformed passerby [57]. Applications for smartglasses address a variety of user needs, from the simple feature of redirecting smartphone notifications [81] to showing maps and street views to aid with navigation [50], video games [69], lifelogging [3,4], and assistive technology for people with visual [2,60,89] and motor impairments [29,36,49].

Just like any other interactive device, smartglasses need some form of input to enable users to react to notifications and to select their preferred choice among the options presented by the user interface. Organizing those options in menus remains one of the most common approach to structure input and to interact effectively with computing systems and devices [11,44], including wearables, and smartglasses are no exception. However, menu design for smartglasses comes with specific challenges, e.g., limited screen real estate to display the menu items and potential occlusion and interference with the user's field of view. Unfortunately, design knowledge to handle such constraints properly is missing and the reasonable option for practitioners is to resort to generic menu designs [10,11,44], which may not always comply with the specific characteristics, form factors, and contexts of use for smartglasses. For example, should menu items be presented on smartglasses horizontally or rather vertically? If horizontally, should menus be displayed at the top, bottom, or in the middle of the see-through lenses and, consequently, in the middle of the central visual field? Are 1-D linear menus preferable to 2-D circular menus in terms of visual appearance? These are just a few examples of possible design options for which little is known about users' preferences.

In this paper, we address these aspects with a design space for smartglasses graphical menus and a vignette experiment [7,30,38] that we conducted to collect, quantify, and analyze users' preferences regarding the visual appearance of smartglasses menus. Our contributions are as follows:

(1) We outline a space of visual design options for graphical menus for smartglasses informed by the literature on graphical menus and guided, among others, by Bertin's visual variables [15] and work from visual aesthetics [37,88].
(2) We conduct a vignette experiment [7,30] with N=251 participants by implementing a randomized A/B testing procedure [72] to quantify preferences regarding the visual appearance of smartglasses menus. To this end, we introduce and employ six measures of *absolute*, *relative*, and *aggregate* preference based on dismissal, preference, and ties rates, the Bradley-Terry model [19] for evaluating design alternatives, and Kendall's coefficient of concordance among raters [40].

MUM 2020, November 22–25, 2020, Essen, Germany

ACM ISBN 978-1-4503-8870-2/20/11…$15.00
https://doi.org/10.1145/3428361.3428467

2 RELATED WORK

We discuss in this section prior work on designing interaction techniques for smartglasses and we briefly overview the vast literature on graphical menus with a focus on visual aesthetics.

2.1 Interaction Techniques for Smartglasses

Interaction techniques for smartglasses generally fall into one of three broad categories: handheld, touch-based either on the device or on the body, and touchless by means of mid-air gesture input and voice commands; see Lee and Hui [45] for a survey and discussion. The question of how to effectively design smartglasses graphical menus remains open since current interaction techniques for menu item selection, such as from [45], have been introduced on the assumption that menu design has already been achieved.

Touch input, in the form of taps and swipe gestures performed on the touchpads embedded in smartglasses can be used for item selection and for scrolling menu items. For example, LYRA [8], a Google Glass application designed to assist flight attendants by displaying situated information regarding requests made by passengers, is controlled with taps and swipes: a tap on the touchpad marks a passenger request as completed, while swipes to the left and right show more information about the passenger. The authors' goal with their design was to keep *"navigation through the application simplistic to allow discrete and easy control"* (p. 213) but essentially this meant falling back on existing design knowledge for interacting with touchscreens and touchpads, repurposed for input on smartglasses. Touch input was also considered by Islam *et al.* [39], who proposed ten tap gestures for an authentication system for smartglasses. Their input method was informed by tapping techniques previously developed for smartwatches, i.e., *"tapping gestures have also been previously explored on wearables... This concept and the timing threshold influenced the design of the tapping gesture set proposed in this paper"* [39, p. 16:4]. Other examples of adaptations of existing techniques to work on smartglasses can be found in the literature; e.g., SwipeZone [35] is a text entry technique for smartglasses that uses gesture input to select letters from a 3×3 rectangular menu, which draws inspiration from input on ultra-small touchscreens [23]; the WISEGlass system [82] leveraged context awareness and implemented input multimodality for contextual menus; and Dingler *et al.* [25] were interested in gesture commands for Rapid Serial Visual Presentation (RSVP) controls to be consistently used across smartphones, smartwatches, and smartglasses. Such examples show how, in the absence of explicit design knowledge on smartglasses graphical menus, the community has resorted to adapting existing designs and interaction techniques to accommodate the form factors and specific characteristics of smartglasses. While successful in some situations, this approach may not work in others, such as touch input on smartglasses for users with upper-body motor impairments [48,49].

2.2 Design Principles for Graphical Menus

A large literature is available on designing graphical menus and corresponding interaction techniques; see [11] for a review and discussion. Graphical menus can be linear, rectangular, circular, horizontal, vertical, static, dynamic, contextual, adaptive and can feature many modalities to present items or to enable users to select those items. For example, *contextual menus* display options that depend on the current task or context of use, limiting the number of items presented to users to those relevant for the task at hand; e.g., after selecting a home appliance, a contextual menu is displayed to show actions available to control that appliance [42]. *Non-contextual menus* are independent of the underlying context and, therefore, can be displayed according to generic, scene-independent criteria. How to display graphical menus on smartglasses, however, introduces new challenges represented by the limited screen real estate, potential occlusion with the visual scene and, in some cases, limited input capabilities of the device itself [45]. Some smartglasses models have relaxed these constraints, e.g., the Epson Moverio BT-350 features a large optical display and Microsoft HoloLens [52] complements 2-D user interface controls, e.g., windows and menus, with photorealistic 3-D graphics aligned with the real world.

The extensive body of work on menu design for immerse virtual environments (VR) could potentially be exploited to inform design for AR smartglasses, including user interfaces and menus, but further validation experiments are necessary. For example, Mulder [53] compared five techniques for selecting items from a 2-D menu projected in a 3-D virtual environment: ray casting, sight-based and orthogonal projective techniques, 3-D positioning, and mouse-based selection. Their results showed that mouse-based selection was the fastest technique, followed by orthogonal, sight-based, ray casting, and 3-D positioning; however, for a large number of menu items, ray casting was recommended [53], which implies that additional equipment must be available to track users' hand movements and pointing gestures. Although wearing additional equipment might not be a major issue in VR, such as for the "Tulip" menu [18], it may prove cumbersome for mobile scenarios.

2.3 Vignette Studies to Collect Perceptions of and Responses to Hypothetical Situations

In this work, we conduct a vignette study [30] in order to collect and examine the perceptions of potential users regarding a variety of graphical menu designs for smartglasses. In the rest of this section, we present what a vignette study is since such studies have been little employed in the HCI research and practice [34,38,46], and we highlight the strengths and weaknesses of these type of studies.

A "vignette" is a description of a hypothetical situation, such as Souza's [24] illustration of the risky decision taken by a naive user to store sensitive information in the internal storage of the web browser running on their computer. Finch [30] described vignettes as *"short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond"* (p. 105). More generally, a vignette is *"a short, carefully constructed description of a person, object, or situation, representing a systematic combination of characteristics"* [7, p. 128]. In the case of smartglasses menus, a vignette represents any textual or graphical depiction suggesting a hypothetical arrangement and display of menu items. However, even though the vignette is a fiction, it nevertheless illustrates a potential real-world situation or scenario and, consequently, vignettes operate in the realm of possibilities for the phenomena and actions they describe. Thus, an important characteristic of vignettes is that they empower interviewees and study participants *"to define the situation [depicted by the vignette] in their* *own terms"* [14]. Finch [30] noted the distinctiveness of the vignette method with respect to other survey techniques: *"vignettes do make possible one particular form of open-ended question which is situationally specific"* (p. 106). According to this perspective, vignettes approach projective techniques from psychology that enable study participants to define the meaning of the situation for themselves. This characteristic of vignettes comes very handy during interviews since it limits the influence of the interviewer on inflicting their perspective onto the interviewee.

When a vignette is employed as part of an experiment, such as applying vignettes in psychology or sociology [13,14,16,30,84] where participants are introduced to the situation pictured by the vignette and their impressions and perceptions of that situation are collected, the experiment is called a "vignette study" [1,7,9,28,59,65]. In such studies, the descriptions delivered by vignettes represent the support for the study participants to form their understanding of the hypothetical situation under investigation and to respond to that situation. Atzmüller and Steiner [7] pointed to the strengths and weaknesses of vignette studies compared to other methods. For example, while traditional surveys have high external validity, they also exhibit low internal validity caused by the multicollinearity of the measured variables; and traditional experiment designs have high internal validity, but may exhibit low external validity because of the samples of participants not being representative of the target population [7]. In this context, vignette studies address the limitations of both surveys and experiments [7,65]: according to Aguinis and Bradley [1], the vignette method represents *"a way to address the dilemma of conducting experimental research that results in high levels of confidence regarding internal validity but is challenged by threats to external validity versus conducting nonexperimental research that usually maximizes external validity but whose conclusions are ambiguous regarding causal relationships"* (p. 351).

The use of vignette studies has been scarce in the HCI community, where we were able to identify just a handful of papers. For example, Lindgaard *et al.* [46] used vignettes to present participants with fictitious cover stories about patient symptoms in their study to inform the design of diagnostic decision support systems; Ellis and Tyre [26] conducted a vignette study to examine help-seeking and help-giving by technology users and technical specialists; and Goodman *et al.* [34] used vignettes to illustrate everyday interaction design work to demonstrate how professionals negotiate research questions relevant to interaction design. More recently, Hoyle *et al.* [38] reported results from an online factorial vignette study conducted using Amazon Mechanical Turk with N=279 participants to collect judgments regarding the appropriateness of posting private photographs online.

Our choice for a vignette experiment to conduct our scientific investigation vs. other types of research methods will be thoroughly motivated in Section 4. For now, we simply note the flexibility of the vignette method for interviewers and interviewees alike as well as the opportunity to administer it over the web, e.g., by delivering it as part of a web-based tool [72], in order to reach a large audience of participants (N=251 in our experiment), difficult to reach otherwise in laboratory studies. We also note the opportunity of web-based vignette experiments to enable HCI researchers to collect responses from participants with reduced mobility, such as people with motor impairments, or people under the constraints of social distancing.

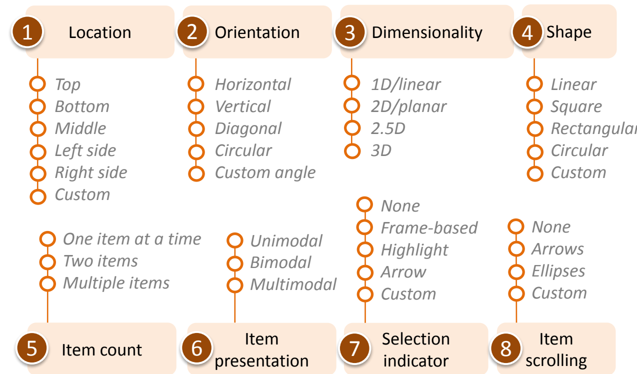

Figure 1: A design space for smartglasses menus guided by Bertin's [15] visual variables, Hashimoto's [37] and Zen and Vanderdonckt's [88] work on visual aesthetics, and by connecting design characteristics with design goals [61].

3 A DESIGN SPACE FOR SMARTGLASSES GRAPHICAL MENUS

We propose a design space to structure a practical problem, *how to design visual menus for smartglasses?*, into a *solution space* that can be readily used to arrive at clear design options. To this end, we build on a rich literature on designing visual menus [10,11,54,62], from which we draw inspiration with respect to input modalities [10,11], input information [54], menu item geometry and scrolling [10], menu geometry, and menu position [62]. Also, since we focus on visual menus, we adopt Bertin's [15] visual variables as guidance to structure the dimensions of our design space by considering position, orientation, size, shape, value, color, texture, and motion options. We equally draw from the work of Hashimoto [37] on the fundamentals of visual design and from Zen and Vanderdonckt [88] on the aesthetics of graphical UIs. For example, according to Hashimoto [37], visual design applied to graphical UIs aims to optimize the usage of visual elements to ensure aesthetics toward a rich user experience. Thus, visual aesthetics can be assessed by evaluating visual properties, such as balance, symmetry, equilibrium, and proportion as part of the multi-factorial aspect of visual design [88]. Lastly, we follow the principles of *"connecting design characteristics with design goals"* of Samp [61], which we adapt to smartglasses graphical menus. For example, we use the visual structure principle to inform the item presentation dimension in our design space and the item shape categories to inform various shapes for smartglasses graphical menus.

Based on the above considerations, our design space consists of eight dimensions: (1) menu location, (2) shape, (3) orientation, (4) dimensionality, (5) item presentation, (6) item count, (7) selection indicator, and (8) item scrolling; see Figure 1. For example, the location of the menu can be at the top, bottom, in the middle, on the left or right side of the lenses; menu items can be displayed horizontally, vertically, diagonally, circularly, or at a custom angle; the shape of the menu can be linear, rectangular, etc. Our design space distills established concepts from the literature on menu design and visual aesthetics into the first adaptation for smartglasses.

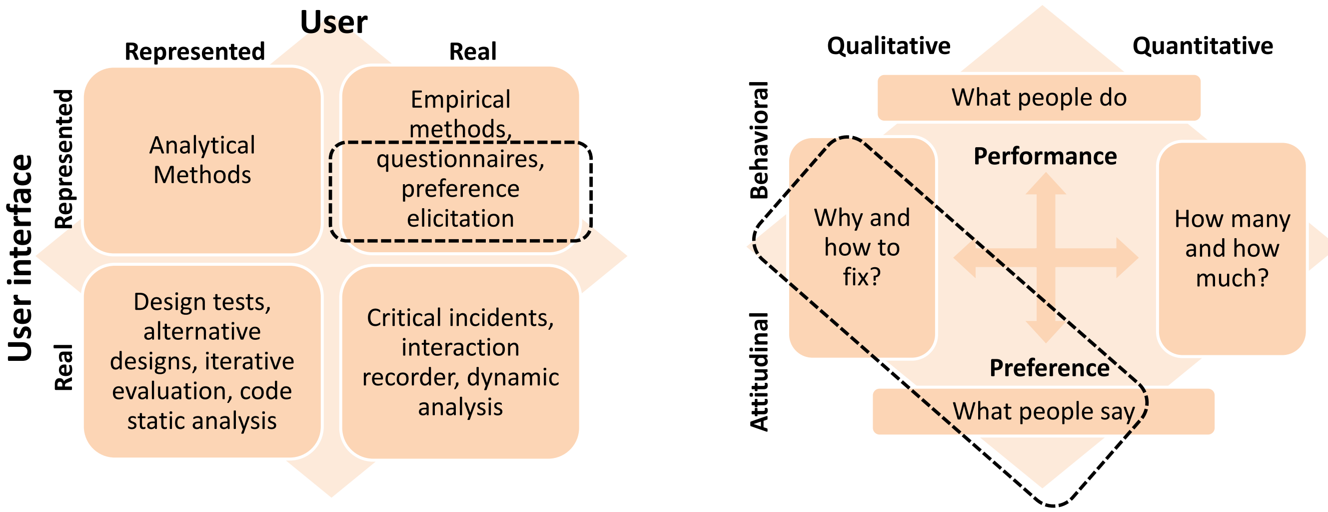

Figure 2: The scope of our experiment in terms of the frameworks of Whitefield *et al.* [83] (left) and Rohrer [58] (right).

4 EXPERIMENT

We conducted a vignette study [7,30,38] to collect prospective users' preferences regarding the visual appearance of smartglasses menus. In our study, a vignette consists of a graphical mock-up of a smartglasses menu derived by manipulating one or multiple dimensions of our design space, e.g., *location* and *item presentation*. Two such vignettes are presented simultaneously to participants to generate a contrasting effect. To control the presentation of multiple pairs, we employed the A/B testing method [41] due to its popularity in many areas of research [12,41,87]. According to this method, participants are elicited for their preferences regarding several alternatives. We used the AB4Web software tool [72] to implement the A/B testing method. Before proceeding further with the description of our experiment design, we briefly present the scope of our vignette study since such studies have been little employed in HCI.

4.1 The Scope of Our Experiment

Since we collect the preferences of prospective (not actual) users regarding mock-up (not actual) designs of smartglasses menus, it is important to clarify the contextual framework in which our experiment is conducted next to its specific nature of a vignette study [7,13,30]. To this end, Figure 2 illustrates the scope of our experiment in terms of two frameworks proposed by Whitefield *et al.* [83] and Rohrer [58], respectively.

The dimensions of Whitefield *et al.* [83] instantiate both real artifacts and representations of the user and the UI. In our experiment, we target prospected users, for which little to no experience in wearing and using smartglasses is expected, in order to reach a sample of participants as large and diverse as possible. Similarly, smartglasses menus are represented in our experiment by means of graphical abstractions (the vignettes) in order not to influence participants with the particular look and feel induced by a specific operating system, model of smartglasses, or application. The top-right quadrant of Figure 2, left reflects the positioning desired for our experiment in terms of the framework from [83].

The two dimensions from Rohrer [58] contrast the qualitative vs. the quantitative (on the horizontal axis) and the attitudinal vs. the behavioral components (on the vertical axis) for empirical research studies. The first contrast establishes whether the research method produces data about participants by means of direct observation as opposed to using indirect observations. The second contrast makes the distinction between perceived attitudes and participants' actual behavior. We believe that capturing participants' behavior indirectly on various smartglasses designs may lead to inconsistencies due to the various form factors of those devices and their heterogeneous interaction capabilities. Consequently, we designed our experiment to collect data directly in terms of preferences for the visual appearance of menus depicted using vignettes in ways that are independent of any smartglasses model, operating system, or application, which positions our experiment design in the bottom-left part of Rohrer's [58] framework; see Figure 2, right.

4.2 Participants

A number of 290 volunteers responded to our invitation to participate in an online questionnaire implemented using AB4Web [72]. Participants' demographics covered a broad range of nationalities from seventeen countries (in alphabetical order: Algeria, Austria, Belgium, Brazil, Columbia, France, Germany, Greece, Japan, Luxembourg, Portugal, Romania, Spain, The Netherlands, Tunisia, UK, USA), spoken languages, and professional domains (e.g., administration, business, education, finance, government, health care, information technology, law, management, software development, and students of various fields of study). We discarded the responses of 39 participants (13.7%) because of incomplete data or uncommitted responses, such as participants repeatedly entering the same answer (e.g., always picking the left or the right variant for all or most of the comparisons) or too rapid responses. In the end, the number of effective participants with valid data considered for our analysis was N=251. Participants' ages ranged between 10 and 81 years old (M=31.1, SD=12.4 years) and 30.7% were female.

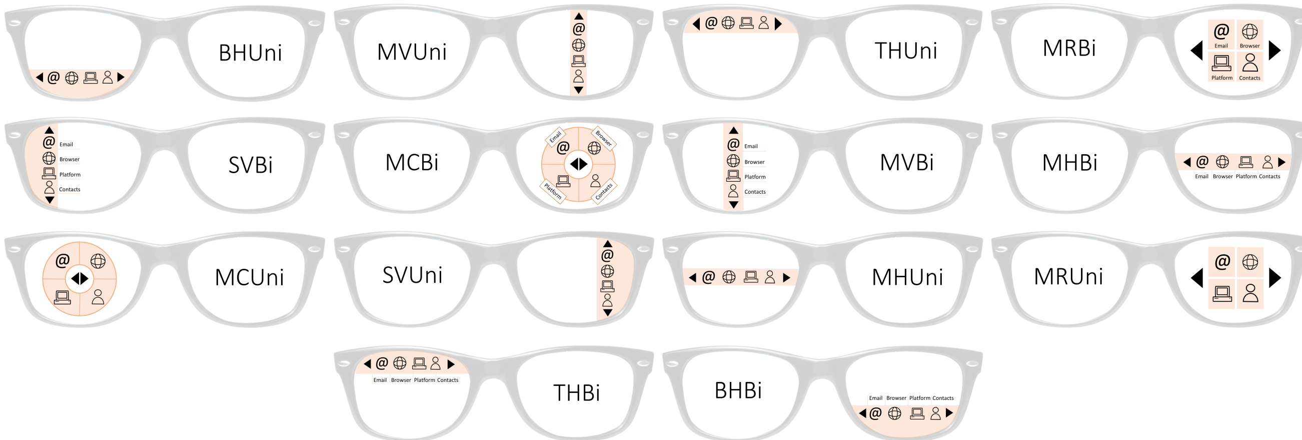

Figure 3: The fourteen vignettes employed in our experiment. *Notes:* each menu is encoded using the notation LOM, where L stands for location (T for top, B for bottom, S for side, M for middle), O for orientation (H horizontal, V vertical, C circular, R rectangular), and M for the modality of presentation (Uni for unimodal icons and Bi for bimodal text and icons).

4.3 Stimuli and Apparatus

Because of the wide range of possible combinations of the options from our design space (Figure 1), we adopted two guiding principles to reduce the number of distinct menu designs for our vignette experiment so that the experiment would be (i) practical in terms of the workload expected from participants to complete the study and (ii) to cover as many design options as possible without affecting the representation quality of the smartglasses menu vignettes. Our two guiding principles address the suitability of graphical menus represented using vignettes and directing participants' attention toward the menu design as a whole:

(1) Smartglasses menus should be effectively presented in the form of 2-D graphical vignettes without any loss in representation quality because of dimensionality reduction or conversion of non-graphical modalities to graphical output. For example, 2.5-D and 3-D menus are less straightforward to represent on a 2-D display compared to 1-D/2-D designs and, thus, we eliminated *dimensionality* from our experiment.
(2) The vignette should not draw participants' attention toward specific menu items, but rather focus it on the visual appearance of the menu design as a whole. Furthermore, depicting individual menu items in specific states (e.g., selected, disabled, etc.) would have considerably increased the number of vignettes to be presented to our participants. Therefore, we decided to eliminate the *selection indicator* and *item scrolling* design dimensions from our experiment. Future work can address users' perceptions regarding item-specific visual appearance, such as evaluating various types of selection indicators to indicate item scrolling options.

By following these two guiding principles, we were left with the *location*, *orientation*, *shape*, *item count*, and *item presentation* dimensions. To eliminate the potential effect of the number of menu items on participants' perceptions of visual appearance, we decided that each vignette should present the same number of menu items. In the end, we selected a number of 14 representative menu types defined by the following four dimensions: *location* (with the options: top, bottom, middle, and side), *orientation* (horizontal and vertical), *shape* (linear and rectangular), and *item presentation* (unimodal icons and bimodal text and icons). We created mock-ups of each of the 14 smartglasses menu types with six menu items: four generic items ("Send email", "Open browser", "Connect glasses to platform", and "Show list of contacts") and two navigation items ("Previous" and "Next"). For each menu design, a vectorial image was produced (see Figure 3) for the AB4Web application [72].

4.4 Task

Participants were presented with pairs of vignettes depicting menu design alternatives and were asked to select the variants they preferred most by relying exclusively on the visual appearance of the presented information. In case participants were undecided, AB4Web [72] offered a "draw" option. By following the open endedness of vignette studies that expect respondents to define the situation depicted by the vignettes by themselves [14,30], we did not provide any specific guidance to our participants regarding how visual appearance should be evaluated nor any context information (i.e., there was no illustration of any background perceivable through the glasses or any suggestion of an application domain). By adopting this minimalistic representation for menus and by not anchoring our participants' preferences in any set of visual aesthetics criteria or specific context of use for smartglasses, we aimed for eliciting a wide range of preferences resulted from self-formed impressions of what visual appearances would work best for menus shown in the context of a pair of glasses. In other types of studies conducted to elicit users' preferences, such as end-user gesture elicitation [73,79,86], similar minimalistic approaches are adopted by removing unnecessary details, such as details of graphical UIs pertaining to some specific operating system that might influence participants' responses. In some instances of applying the end-user elicitation method [33] or in the case of standard forms of questionnaires and surveys that elicit users' preferences [56], participants are simply asked to imagine the effect based on a textual description of the situation under examination. In our case, the use of vignettes assures representations of graphical menus independent of any smartglasses model, operating system, or application. Note that this approach works well when the expected participation is large (the goal of our study) in order to understand consensus formation from many responses and might not work as well for in-the-lab studies with few participants, for which controlled conditions are either explicit (e.g., participants are presented with an actual graphical representation of a menu) or implicit (e.g., there is an actual pair of smartglasses of a given brand and model). Also, we followed recommendations for implementing vignette studies, e.g., the presented stories should not be too complex, should contain sufficient context for participants to form an understanding about the depicted situation but also be vague enough to force them to contribute factors that influence their decisions, and stories must appear plausible, real, and reflect mundane occurrences [14]. These recommendations are reflected in the context-independent, device-independent, application-independent, and operating system-independent graphical representations of the menus illustrated in Figure 3.

A simulation revealed that about 30 minutes were needed to complete all the $14 \cdot (14-1)/2 = 91$ trials, which we considered a duration too long for volunteers of an online questionnaire with potential negative influence on the completion rate of the experiment and the quality of the results because of fatigue or boredom effects. Therefore, a restricted number of 30 pairs of vignettes were randomly generated for each participant making sure that, overall, the numbers of presentations of each menu type, across all the participants of the study, were roughly equal. The order of pairs was randomized across participants and the order of presenting the menus in each pair (left *vs.* right) was randomized across pairs. No time constraint was imposed. With this procedure, participants needed between 5 and 10 minutes to complete the questionnaire.

4.5 Measures

Computing consensus among users can be accomplished in several ways, such as by employing Condorcet, Borda, or Dowdall counts [27], Kendall's coefficient of concordance [40], Bradley-Terry's estimates of abilities [19], or agreement and coagreement measures from end-user elicitation studies [77,78,86]. For example, the Borda count is a consensus-based voting system that determines the outcome of a debate by assigning to each candidate a number of points corresponding to the number of candidates that were ranked lower. Another popular method was introduced by Bradley and Terry [19], who proposed a model for estimating the probability of preference for a specific element among many options. The Bradley-Terry model has been adopted in the practice of HCI, such as by Serrano *et al.* [63] to analyze user preferences for visual designs of rectangular displays; Al Maimani and Roudaut [5] employed it to examine users' preferences for haptic feedback; Takao *et al.* [68] used the model to understand the effects of differences in timing between motion and utterance; and Chen *et al.* [22] evaluated the Quality of Experience of multimedia content using crowdsourcing. Several measures of agreement [6,74,77,78,86] have been popularized in the context of gesture elicitation studies [80,85,86] to inform, among other, the selection of "winning" gestures representative of users' preferences for specific devices or specific contexts of use. Informed by this large literature on computing consensus, we decided to employ multiple measures instead of just one; see next. But first, we introduce two concepts:

(1) Num-Presentations represents the number of times that a particular menu was presented to participants as part of the A/B-type comparisons. For example, if the BHBi design was randomly selected to be presented in seven distinct A/B-type comparisons for participant P_{12}, then Num-Presentations is 7. However, Num-Presentations can be aggregated for the whole experiment, resulting in the total amount of presentations of a specific menu, e.g., BHBi was presented 1,095 times, all 251 participants considered.
(2) Preference-Matrix is a symmetric N×N matrix for which cells store the number of preference votes for menu m_i when compared to m_j. For example, if menu BHBi was preferred 79 times in direct comparisons with BHUni and other 25 comparisons were ties, then the corresponding cell of the Preference-Matrix contains $79 \cdot 1 + 25 \cdot 0.5 = 91.5$ votes.

Based on these considerations, our measures to quantify users' preferences for the visual appearance of menus are as follows:

(1) **Absolute measures of preference (reported per menu):**
  (a) Preference-Rate represents the percentage of positive votes received by a specific menu, all participants considered; e.g., if the BHBi menu design was preferred 551 times over other menus from a total number of 1,095 comparisons, then its Preference-Rate is $551/1095 = 50.3\%$.
  (b) Dismissal-Rate represents the percentage of negative votes received by a specific menu type, all participants considered; e.g., if the BHBi menu design from the previous example was not the preferred option during 374 of the total number of comparisons, then its Dismissal-Rate is $374/1095 = 34.2\%$.
  (c) Ties-Rate represents the percentage of tie votes received by a specific menu type, defined as 100 - (Preference-Rate + Dismissal-Rate). For our example, Ties-Rate is 100 - (50.3 + 34.2) = 15.5%.
  (d) Latent-Preference-Rate is the difference between the preference and dismissal percentages for a given menu type; e.g., the latent preference for the BHBi menu design is (551-374)/1095=16.2%. Note that Latent-Preference-Rate can take negative values as well between -100% (absolute dismissal) and 100% (absolute preference).
(2) **Measures of relative preference (for pairs of menus):**
  (e) Relative-Preference ($\rho_{i,j}$) represents the probability that menu m_i will be preferred to m_j. We used the Bradley-Terry model [19] to estimate the probability that a menu design is preferable over the rest. The model works by assigning strength parameters π_i to each menu (also called "abilities"), based on the idea that competition results inform the underlying abilities of the competitors, which are menu designs in our case. The abilities π_i provide a ranking of the menu designs, but also the relative preference probabilities for each menu m_i, as follows:

$$\rho_{i,j} = \frac{\pi_i}{\pi_i + \pi_j} = \frac{1}{1 + e^{-(log(\pi_i) - log(\pi_j))}} \quad (1)$$

where the first expression is easier to compute for our purposes and the second is suited for logistic modeling [70].

(3) **Measures of aggregate preference (reported for the entire set of menus):**

(f) CONCORDANCE. We used Kendall's [40] W coefficient of concordance to compute the agreement among participants in terms of their rankings of the preferred menus. The W coefficient ranges from 0 (no agreement, no overall trend) to 1 (absolute agreement, each participant has the same order of preferences, or ranking, for smartglasses menus). To compute W, we computed a ranking of the menus for each participant using a Borda count method starting from 1; see [27] for a description. Specifically, we assigned 3 points to each menu design when that menu won a direct A/B-type comparison, 2 points for each draw, and 1 point when the alternative menu was preferred. For example, if the BHBi menu design was preferred 551 times, dismissed 374 times, and ended up in a draw for 170 direct comparisons with other menu designs, its Borda count was 3·551+2·170+1·374=2,367.

5 RESULTS

We report in this section results on the *absolute*, *relative*, and *aggregate* preference regarding the visual appearance of smartglasses menus. But first, we discuss the representativeness of our sample of participants and the validity of the randomized A/B procedure.

5.1 Data Validity and Representativeness

The data collected from our study, after removing the uncommitted participants (see the previous section), are represented by 7,530 preference responses collected from 251 (participants) × 30 (randomized A/B trials) regarding the visual appearance of 14 smartglasses menus designed with the *shape*, *orientation*, *location*, and *item presentation* dimensions from our design space. In the following, we analyze the age and gender distribution of our participants and the random generation of the A/B pairs to confirm (1) the representativeness of our sample for the target user population and (2) the validity of our randomized A/B procedure in terms of the number of presentations of each vignette.

5.2 Representativeness of Respondents

Figure 4 illustrates the age-gender demographic distribution of our participants showing a good age coverage for both gender groups: female participants between 11 and 63 years old and males between 10 and 81 years. The figure also shows that age distributions are not normal (an observation confirmed by Shapiro-Wilk tests, W=.921, p<.001 for female and W=.832, p<.001 for male participants, respectively) with a higher representativeness of young people, less than 35 years old, in our sample. This outcome is fortunate, however, as it is a known fact that young people use a greater breadth of technologies compared to older adults and are more open to new technology, whereas older adults are more likely to use technologies that have been around for a longer period of time [55]. From this perspective, the population sample fits well the topic of our investigation, which is focused on understanding preferences for a new type of smart gadget made available only recently to

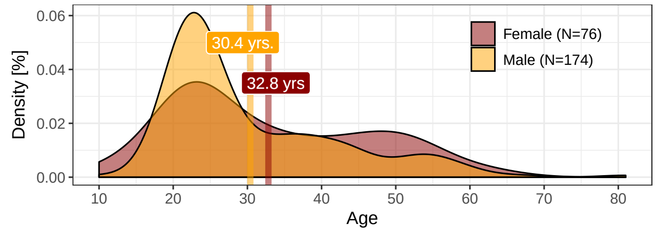

Figure 4: Statistics for our participants' age-gender demographics with mean ages highlighted.

the large public. Moreover, the mean ages are close for the two gender groups (M=32.8, Mdn=29.0 years for female and M=30.4, Mdn=25.0 years for male participants, respectively, Wilcoxon's rank sum W=7367, p=.207, $n.s.$) and the age distributions are not significantly different (as indicated by a Kolmogorov-Smirnov test D=.147, p=.175, $n.s.$). Although non significant results cannot be used to accept null hypotheses, they increase our confidence in the representativeness of our sample of participants. To summarize, we report in this study users' preferences (i) across a large variety of age groups with (ii) similar age distributions for men and women, while (iii) the opinions of the age group most likely to adopt and use smartglasses, i.e., young people less than 35 years old, are given appropriate representativeness in our sample.

5.3 Validity of the A/B Randomized Procedure

Table 1 lists the number of presentations of each menu design across all the randomized A/B trials, e.g., the BHBi design was presented 1,095 times, all participants considered, out of all the 7,530 (A/B-type trials) × 2 (vignettes presented per trial) = 15,060 total number of menu presentations. The expected number of presentations per menu is 15,060 divided by 14 (distinct menu designs) = 1,075. The standard deviation was 16.98 with absolute errors varying between 0 and 32 (0% and 2.98%) with respect to the expected mean. These results confirm that the randomized A/B procedure was applied correctly with no menu design being favored over the others in terms of a significantly larger presentation count.

Table 1: Characteristics and number of presentations of each of our fourteen vignettes representing smartglasses menu designs. Overall, a total number of 15,060 vignettes were presented to N=251 participants.

Menu design	Location	Orientation & Shape	Item presentation	Preference count	Presentation count
BHBi	bottom	horizontal	bimodal	551	1095
BHUni	bottom	horizontal	unimodal	698	1107
MCBi	middle	circular	bimodal	274	1075
MCUni	middle	circular	unimodal	370	1063
MHBi	middle	horizontal	bimodal	354	1064
MHUni	middle	horizontal	unimodal	434	1068
MRBi	middle	rectangular	bimodal	352	1065
MRUni	middle	rectangular	unimodal	308	1101
MVBi	middle	vertical	bimodal	455	1099
MVUni	middle	vertical	unimodal	217	1067
SVBi	side	vertical	bimodal	727	1072
SVUni	side	vertical	unimodal	477	1062
THBi	top	horizontal	bimodal	693	1065
THUni	top	horizontal	unimodal	685	1057
				Total	**15,060**

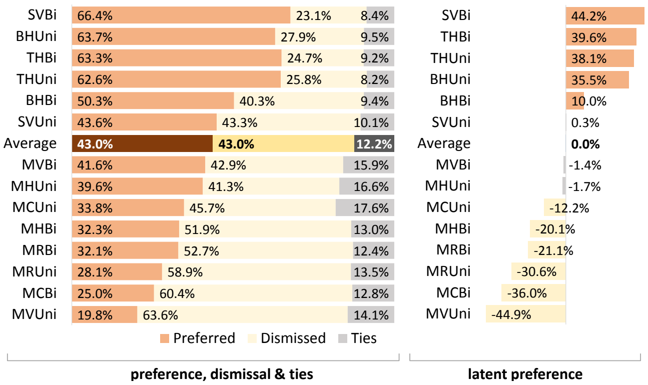

Figure 5: Preferences expressed for our menu designs.

5.4 Absolute Preferences

Figure 5 shows participants' preferences for our set of 14 menu designs by reporting the PREFERENCE-RATE, DISMISSAL-RATE, TIES-RATE (Figure 5, left), and LATENT-PREFERENCE-RATE measures (Figure 5, right). The most preferred menu was SVBi (i.e., side location, vertical arrangement of items, and bimodal presentation using both text and icons) with a preference rate of 66.4%, 8.4% ties, and a latent preference of 44.2%. This result suggests that when menu items are presented on the side of the smartglasses to address peripheral vision, they should be aligned vertically to optimize navigation and visual search, while text descriptions should be presented next to the corresponding icons to assist exploration. The next three positions were occupied by the BHUni, THBi, and THUni menus with 63.7%, 63.3%, and 62.6% preference rates and 35.5%, 39.6%, and 38.1% latent preference rates, respectively. These findings show that if menu items are presented at the bottom or the top of the lenses, they should be presented horizontally and using icons only. The least preferred menu was MVUni, probably because of its middle location, likely to affect overall visibility, vertically arranged items being more difficult to scan, and taking up too much space with the bimodal presentation.

Figure 6 reports preference rates for each applicable design dimension, which we computed by counting the number of positive votes (+1) and half a vote for ties (+0.5) each time a specific menu was preferred to others. For instance, each time the BHBi menu (bottom, horizontal, bimodal) won a direct comparison, the bottom location received 1 point, the horizontal orientation received 1 point, and the bimodal modality of item presentation received 1 point, respectively. Similarly, each time the BHBi design ended up in a tie, each of its design options received 0.5 points. By dividing the result to the total number of presentations of each menu, we obtained the preference rates from Figure 6. For example, the bottom location received 1,352.5 points from 2,202 direct comparisons (involving BHBi and BHUni) and, thus, its preference rate was $1352.5/2202 \cdot 100\% = 61.4\%$. Also, the middle location received more points (3,398.5) from more comparisons (8,602) due to the MCBi, MCUni, MHBi, MHUni, MRBi, MRUni, MVBi, and MVUni designs, and resulted in a lower preference rate compared to the bottom location, $3398.5/8602 \cdot 100\% = 39.5\%$. Regarding the *orientation* and *shape* dimensions, the linear horizontal design was preferred in 58.5% of the cases, followed by linear vertical (49.8%), while the circular and rectangular designs both scored 37%. The distribution between unimodal and bimodal designs was roughly equal, with slightly more preference for the bimodal presentation of menu items using both text and icons (51.1% *vs.* 48.9%).

5.5 Relative Preferences

Besides the absolute preference rates, we were also interested in relative preferences, e.g., how likely is it for menu m_i to be preferred over m_j. To this end, we fit the Bradley-Terry model to our data using the BradleyTerry2 R library [70], which uses the glm.fit method with the binomial family and the logit link function. The model needed four Fisher scoring iterations to converge, and the analysis indicated a better fit (AIC=397.88, residual deviance 105.73, df=64) compared to an intercept-only model (AIC=1088, null deviance 1088, df=77).

Figure 7 reports the ability π_i of each menu m_i to win users' preferences computed by the Bradley-Terry model [70]. These results confirmed our previous findings (see Figure 5): SVBi had the greatest ability (0.75), followed by THBi (0.67), THUni (0.61), and BHUni (0.55). We found no significant differences between SVBi and THBi (Z=−0.894, p=.371, $n.s.$) nor between SVBi and THUni (Z=−1.494, p=.135, $n.s.$), but the SVBi menu design had significantly larger ability than all the other menus (p<.05 for BHUni and p<.001 for the other menus) to attract participants' preferences. Based on the values of abilities π_i and Eq. 1, Figure 8 illustrates the probabilities $\rho_{i,j}$ that menu design m_i will be preferred over menu m_j, e.g., SVBi (side location, vertical orientation, bimodal presentation) is likely to be preferred over MCUni (middle, circular, unimodal) in a direct comparison with 77.4% chance.

5.6 Aggregate Preference

We employed Kendall's [40] W coefficient of concordance to understand the consensus between participants' rankings of menus, which we computed using the Borda count [27]. We found W=0.109 ($\chi^2_{(250)}$=356.32, p<.001), a result that indicates low consensus and a diversity of rankings for the set of fourteen menus. Our previous analysis of cumulated preferences showed that consensus over particular menu designs emerged nonetheless despite a wide variety of individual rankings; see Figure 5. However, this result suggests that groups of participants with similar preferences in terms of menu rankings could potentially be identified by running further exploratory analysis. For example, we found that male participants were slightly more in agreement about their menu rankings than female participants (W=0.120 *vs.* W=.094, respectively, both $p < .001$).

6 DISCUSSION

We present in this section a summary of our results, which we distill into takeaways. We also discuss limitations of our experiment and suggest ways to address those limitations in future work.

6.1 Summary of Results and Takeaways

Our empirical results showed *location* to be the most determinant variable for smartglasses graphical menus with some design options for *location* winning participants' preferences in over 60%

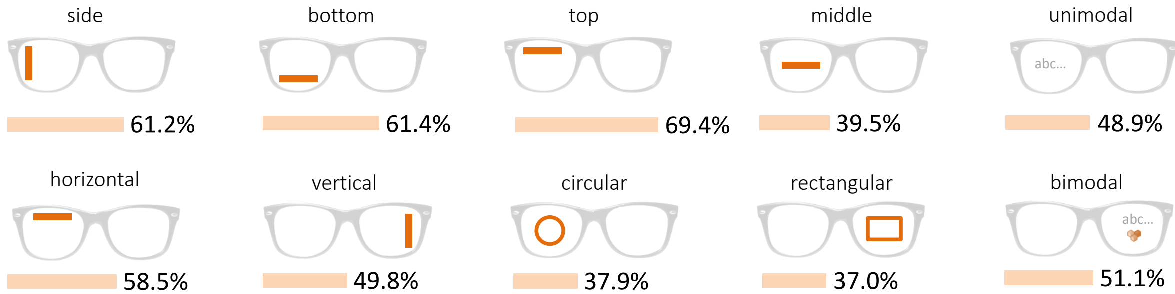

Figure 6: Preference rates for various design options regarding *menu location, orientation, shape,* and *item presentation.*

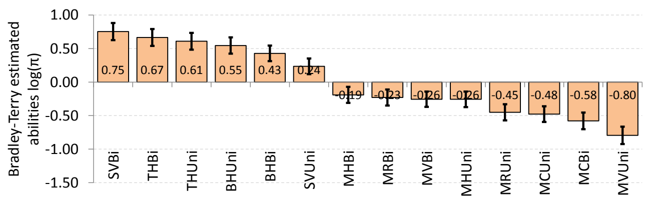

Figure 7: Bradley-Terry abilities estimated for each menu.

	SVBi	THBi	THUni	BHUni	BHBi	SVUni	MHBi	MRBi	MVBi	MHUni	MRUni	MCUni	MCBi	MVUni
SVBi		52.2%	53.6%	55.2%	58.1%	62.7%	72.0%	72.8%	73.3%	73.3%	76.9%	77.4%	79.1%	82.5%
THBi	47.8%		51.4%	53.0%	55.9%	60.6%	70.2%	71.0%	71.6%	71.6%	75.3%	75.8%	77.6%	81.2%
THUni	46.4%	48.6%		51.6%	54.5%	59.2%	69.0%	69.8%	70.4%	70.4%	74.3%	74.8%	76.6%	80.3%
BHUni	44.8%	47.0%	48.4%		52.9%	57.7%	67.6%	68.5%	69.0%	69.1%	73.0%	73.6%	75.5%	79.3%
BHBi	41.9%	44.1%	45.5%	47.1%		54.8%	65.0%	65.9%	66.5%	66.5%	70.7%	71.2%	73.2%	77.3%
SVUni	37.3%	39.4%	40.8%	42.3%	45.2%		60.5%	61.4%	62.0%	62.1%	66.5%	67.1%	69.3%	73.7%
MHBi	28.0%	29.8%	31.0%	32.4%	35.0%	39.5%		51.0%	51.6%	51.7%	56.5%	57.1%	59.6%	64.6%
MRBi	27.2%	29.0%	30.2%	31.5%	34.1%	38.6%	49.0%		50.7%	50.7%	55.5%	56.2%	58.6%	63.8%
MVBi	26.7%	28.4%	29.6%	31.0%	33.5%	38.0%	48.4%	49.3%		50.0%	54.8%	55.5%	58.0%	63.1%
MHUni	26.7%	28.4%	29.6%	30.9%	33.5%	37.9%	48.3%	49.3%	50.0%		54.8%	55.5%	58.0%	63.1%
MRUni	23.1%	24.7%	25.7%	27.0%	29.3%	33.5%	43.5%	44.5%	45.2%	45.2%		50.7%	53.2%	58.5%
MCUni	22.6%	24.2%	25.2%	26.4%	28.8%	32.9%	42.9%	43.8%	44.5%	44.5%	49.3%		52.5%	57.9%
MCBi	20.9%	22.4%	23.4%	24.5%	26.8%	30.7%	40.4%	41.4%	42.0%	42.0%	46.8%	47.5%		55.4%
MVUni	17.5%	18.8%	19.7%	20.7%	22.7%	26.3%	35.4%	36.2%	36.9%	36.9%	41.5%	42.1%	44.6%	

Figure 8: Relative preference probabilities computed from the abilities estimated by the Bradley-Terry model [19,70].

of the direct A/B comparisons. Next followed *orientation*, *shape*, and *item presentation*. In the following, we summarize takeaways, implications, and ideas for future work:

6.1.1 The location of the visual menus should be chosen to minimize interference with the visual scene, achieving thus good visual search and good utilization of the screen real estate. Our results showed that the top location was the most preferred option (69.1% votes), followed by bottom (61.4%) and side (61.2%). These findings suggest interesting directions for future work in terms of positioning content, including menu items and other controls, for smartglasses UIs. For example, the bottom location is normally reserved for proximal vision, useful for reading, while the top location can be assimilated with far-away vision. Consequently, contextual menus could be displayed either at the top or the bottom, depending on the task at hand, e.g., reading or performing visual search. Also, an interesting fact about side menus (61.2% preference votes) is that they do not require the user to shift the focus of attention to the menu and then back to the visual scene, as menu perception can be achieved with the peripheral vision. Our participants' preferences for side menus is a finding in line with performance results from prior work that reported users spending 50% less time looking at the side menu on their smartglasses by exploiting peripheral vision [21].

6.1.2 Menus should not be displayed in the middle. Our results showed that the middle location was the least preferred among all the design options for the *location* dimension (39.5%; see Figure 6), while all the menu designs that featured this option scored preference rates below average (less than 43%; see Figure 5, left) and negative latent preference rates (between -1.4% for the MVBi design and -44.9% for MVUni; see Figure 5, right).

6.1.3 The arrangement and orientation of menu items should optimize visual scanning. The vertical orientation for presenting menu items is not only convenient for the side location (i.e., the SVBi and SVUni designs ranked first and sixth place, respectively; see Figure 5), but is also efficient for visual scanning. When items are left justified, there is no need to reacquire item locations; when items are located and arranged in ways that are consistent across existing applications, they match the common locations where users normally expect menus to be positioned, facilitating thus visual search. Overall, the horizontal orientation was preferred for 58.5% of the menu designs located at the bottom (61.4%) and the top (69.4%) of the lenses; see Figure 6.

6.1.4 Menu designs with 2-D navigation were often dismissed in favor of linearly structured menus. Both circular and rectangular menus were the least preferred (37.9% and 37.0%; see Figure 6). A possible explanation is that these designs suggested a spatial navigation scheme that was perceived more difficult to perform compared to the more simple, linear structure of the horizontal and vertical menu designs. Also, both circular and rectangular shapes require more screen real estate to display compared to linear menus. However, 2-D menus may be useful to depict commonly accepted 2-D orderings of items, such as the digits of 3×3 numerical keyboards. Further studies are needed to understand the compromise between preference and performance for such designs.

6.1.5 Unimodal and bimodal presentations of menu items are equally acceptable. Our empirical results revealed little difference between the unimodal (graphical icons only) and bimodal (text and graphical icons) presentation of menu items (48.9% and 51.1% preference votes, respectively; see Figure 6). This finding suggests new design ideas, such as a smooth transition from bimodal to unimodal menus could also be acceptable to users, while saving screen space and fostering learnability. For example, when menus are first displayed, their presentation could be bimodal, after which menus could progressively switch to a unimodal graphical presentation to free screen space for expert users.

6.2 Limitations

We conducted in this work a large-scale user study (N=251 participants) by taking advantage of a vignette-based experiment design, a web-based tool and online participation, a randomized A/B testing procedure [72], and simple instructions provided to participants regarding the evaluation of the visual appearance of smartglasses menu designs. We managed to foster large participation and commitment from our participants to complete the tasks, but we also acknowledge several limitations of our experiment, as follows:

(1) To keep the user study practical in terms of the workload expected from participants, we defined and applied two guiding principles to reduce the number of dimensions from our design space that we utilized to generate the menu design options and the corresponding vignettes. The dimensions that were not considered were *item scrolling*, *item count*, *selection indicator*, and *dimensionality*. Future work is recommended to examine each dimension in depth to uncover more findings about users' preferences.
(2) In accordance with the specific nature of vignette studies, we did not influence participants in any way regarding their possible interpretations of the vignettes. Instead, we let participants define the parameters of visual appearance regarding smartglasses menus that mattered and made sense the most for them; see Barter and Renold [14] and Finch [30] for the benefits of having respondents defining the situation depicted by the vignettes by themselves as well as our discussion from Section 4 in this regard. Future user studies, such as in-the-lab controlled studies, could measure objective dependent variables regarding the visual aesthetics of graphical UIs, such as variables informed by the metrics employed in [88]. We believe that correlating the values of such objective variables with the self-defined notion of visual appearance will be interesting to explore and will lead to new discoveries for smartglasses graphical menus.
(3) Since our experiment was a vignette study, we used graphical representations of possible designs for smartglasses menus instead of actual implementations on a specific device. This approach brought us several advantages, such as the possibility to conduct evaluations of visual appearance without any influence from specific operating systems or models of smartglasses that might have biased our participants' responses as well as to reach a wide range of participants of different age, cultural backgrounds and professional fields. However, generalization of our findings to actual implementations of smartglasses menus should be done with care. Instead, our empirical results should be seen as the first set of data, collected using a large and diverse sample of potential end users, on which to base further explorations.

In the next section we suggest ways to address these limitations with several ideas for future work regarding graphical menu design for smartglasses.

7 CONCLUSION AND FUTURE WORK

We reported in this paper preference results collected from a vignette experiment with large participation conducted to compare various design options for smartglasses graphical menus. These results represent the first attempt in the community to structure design knowledge for visual menus and smartglasses. Consequently, it is our hope that our work will generate constructive discussion and lead to more studies on smartglasses menus toward accumulating practical design knowledge in this regard.

Several future work directions can be envisaged. For example, we employed only some of Bertin's [15] visual variables to keep the design space manageable for practitioners. Exploration of other variables are recommended, such as *color* (e.g., showing menus with color saturation and/or brightness levels in correspondence with the visual scene); more *orientations* (e.g., items displayed horizontally and vertically) and *shapes* (pie, semi-circular, etc.); *distributed menus*, for which inspiration can be drawn from designing distributed UIs [51] (e.g., splitting menu items across the two lenses or displaying one menu on the left lens and sub-menus on the right one); *adaptive menus* [17,31] in which menu items are displayed according to their frequency of use; or *animated menus* [43], including visual effects, e.g., a smartglasses menu that fades into view. In our experiment, we considered just four of the eight dimensions of our design space in order to keep the experiment manageable and foster commitment from our participants to complete the task. Future work can focus on each dimension individually, including the dimensions that we did not consider in this study, such as *item scrolling* or *selection indicator*, to complete our understanding about users' preferences of the visual appearance of smartglasses menus. Studies addressing the preferences of participants with visual impairments [60] regarding the display of menu items on smartglasses as well as of participants with upper-body motor impairments [48,49] for controlling smartglasses menus, including via touchscreens on mobile devices [71,75,76], wearables [32], and chairables [20], are equally recommended for examination in future work. While the explorations of such directions will definitely contribute important information in the community, our design space can already be employed to assist practitioners in their designs of graphical menus for smartglasses.

ACKNOWLEDGMENTS

This work was supported by a grant of the Romanian Ministry of Education and Research, CCCDI-UEFISCDI, project number PN-III-P2-2.1-PED-2019-0352 (276PED/2020), within PNCDI III.

REFERENCES

[1] Herman Aguinis and Kyle J. Bradley. 2014. Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods* 17, 4 (2014), 351–371. https://doi.org/10.1177%2F1094428114547952
[2] Adrian Aiordăchioae, Ovidiu-Andrei Schipor, and Radu-Daniel Vatavu. 2020. An Inventory of Voice Input Commands for Users with Visual Impairments and Assistive Smartglasses Applications. In *Proceedings of the 15th International Conference on Development and Application Systems (DAS '20)*. IEEE, 146–150. https://doi.org/10.1109/DAS49615.2020.9108915
[3] Adrian Aiordăchioae and Radu-Daniel Vatavu. 2019. Life-Tags: A Smartglasses-Based System for Recording and Abstracting Life with Tag Clouds. *Proceedings of the ACM on Human-Computer Interaction* 3, EICS, Article 15 (June 2019), 22 pages. https://doi.org/10.1145/3331157

[4] Adrian Aiordăchioae, Radu-Daniel Vatavu, and Dorin-Mircea Popovici. 2019. A Design Space for Vehicular Lifelogging to Support Creation of Digital Content in Connected Cars. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Valencia, Spain) *(EICS '19)*. ACM, New York, NY, USA, Article 9, 6 pages. https://doi.org/10.1145/3319499.3328234

[5] Ahmed Al Maimani and Anne Roudaut. 2017. Frozen Suit: Designing a Changeable Stiffness Suit and Its Application to Haptic Games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2440–2448. http://doi.acm.org/10.1145/3025453.3025655

[6] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2018. Crowdsourcing Similarity Judgments for Agreement Analysis in End-User Elicitation Studies. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. ACM, New York, NY, USA, 177–188. https://doi.org/10.1145/3242587.3242621

[7] Christiane Atzmüller and Peter M. Steiner. 2010. Experimental Vignette Studies in Survey Research. *Methodology* 6 (2010), 128–138. https://doi.org/10.1027/1614-2241/a000014

[8] Jonas Auda, Matthias Hoppe, Orkhan Amiraslanov, Bo Zhou, Pascal Knierim, Stefan Schneegass, Albrecht Schmidt, and Paul Lukowicz. 2018. LYRA: Smart Wearable In-flight Service Assistant. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers* (Singapore, Singapore) *(ISWC '18)*. ACM, New York, NY, USA, 212–213. http://doi.acm.org/10.1145/3267242.3267282

[9] Katrin Auspurg and Thomas Hinz. 2014. *Factorial Survey Experiments*. Vol. 175. SAGE Publications, Inc., Thousand Oaks, CA. https://us.sagepub.com/en-us/nam/factorial-survey-experiments/book240309

[10] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2009. MenUA: A Design Space of Menu Techniques. www.gillesbailly.fr/menua/l.

[11] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2016. Visual Menu Techniques. *Comput. Surveys* 49, 4, Article 60 (Dec. 2016), 41 pages. http://doi.acm.org/10.1145/3002171

[12] Eytan Bakshy, Dean Eckles, and Michael S. Bernstein. 2014. Designing and Deploying Online Field Experiments. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. ACM, New York, NY, USA, 283–292. http://doi.acm.org/10.1145/2566486.2567967

[13] Davide Barrera and Vincent Buskens. 2007. Imitation and Learning under Uncertainty: A Vignette Experiment. *International Sociology* 22, 3 (2007), 367–396. https://doi.org/10.1177%2F0268580907076576

[14] Christine Barter and Emma Renold. 1999. The use of vignettes in qualitative research. *Social Research Update* 25 (1999). http://sru.soc.surrey.ac.uk/SRU25.html

[15] Jacques Bertin. 1967. *Sémiologie graphique.* Mouton/Gauthier-Villars, Paris.

[16] M.H. Birnbaum. 1999. How to show that 9>221: Collect judgments in a between-subjects design. *Psychological Methods* 4, 3 (1999), 243–249. https://doi.org/10.1037/1082-989X.4.3.243

[17] Sara Bouzit, Gaëlle Calvary, Denis Chêne, and Jean Vanderdonckt. 2016. A Design Space for Engineering Graphical Adaptive Menus. In *Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '16)*. ACM, New York, NY, USA, 239–244. http://doi.acm.org/10.1145/2933242.2935874

[18] Doug A. Bowman and Chadwick A. Wingrave. 2001. Design and Evaluation of Menu Systems for Immersive Virtual Environments. In *Proceedings of the Virtual Reality 2001 Conference (VR '01)*. IEEE Computer Society, Washington, DC, USA, 149–149. http://dl.acm.org/citation.cfm?id=580521.835855

[19] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952), 324–345. http://www.jstor.org/stable/2334029

[20] Patrick Carrington. 2015. Developing a Chairable Computing Platform to Support Power Wheelchair Users. *SIGACCESS Access. Comput.* 111 (July 2015), 30–33. https://doi.org/10.1145/2809904.2809911

[21] Isha Chaturvedi, Farshid Hassani Bijarbooneh, Tristan Braud, and Pan Hui. 2019. Peripheral Vision: A New Killer App for Smart Glasses. In *Proceedings of the 2019 ACM International Conference on Intelligent User Interfaces (IUI '19)*. ACM, New York, NY, USA, 626–636. https://doi.org/10.1145/3301275.3302263

[22] Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. 2009. A Crowdsourceable QoE Evaluation Framework for Multimedia Content. In *Proceedings of the 17th ACM International Conference on Multimedia (MM '09)*. ACM, New York, NY, USA, 491–500. http://doi.acm.org/10.1145/1631272.1631339

[23] Xiang 'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: A Text Entry Technique for Ultra-small Interfaces That Supports Novice to Expert Transitions. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. ACM, New York, NY, USA, 615–620. http://doi.acm.org/10.1145/2642918.2647354

[24] Clarisse Sieckenius de Souza. 2018. Semiotics and Human-Computer Interaction. In *The Wiley Handbook of Human Computer Interaction*, Kent L. Norman and Jurek Kirakowski (Eds.). John Wiley & Sons Ltd., West Sussex, UK, 33–49. https://doi.org/10.1002/9781118976005.ch2

[25] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 419, 12 pages. https://doi.org/10.1145/3173574.3173993

[26] S. Ellis and M.J. Tyre. 2001. Helping relations between technology users and developers: a vignette study. *IEEE Transactions on Engineering Management* 48, 1 (February 2001), 56–69. https://doi.org/10.1109/17.913166

[27] Peter Emerson. 2013. The original Borda count and partial voting. *Social Choice and Welfare* 40, 2 (2013), 353–358. https://doi.org/10.1007/s00355-011-0603-9

[28] Ines Engelmann. 2017. Vignette Research. In *The International Encyclopedia of Communication Research Methods*, J. Matthes, C.S. Davis, and R.F. Potter (Eds.). John Wiley & Sons, Inc., 1–5. https://doi.org/10.1002/9781118901731.iecrm0263

[29] Andrea Ferracani, Marco Faustino, Gabriele Xavier Giannini, Lea Landucci, and Alberto Del Bimbo. 2017. Natural Experiences in Museums through Virtual Reality and Voice Commands. In *Proceedings of the 25th ACM International Conference on Multimedia* (Mountain View, California, USA) *(MM '17)*. ACM, New York, NY, USA, 1233–1234. https://doi.org/10.1145/3123266.3127916

[30] J. Finch. 1987. The Vignette Technique in Survey Research. *Sociology* 21 (1987), 105–114. https://doi.org/10.1177%2F0038038587021001008

[31] Leah Findlater and Joanna McGrenere. 2004. A Comparison of Static, Adaptive, and Adaptable Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 89–96. http://doi.acm.org/10.1145/985692.985704

[32] Bogdan-Florin Gheran, Ovidiu-Ciprian Ungurean, and Radu-Daniel Vatavu. 2018. Toward Smart Rings as Assistive Devices for People with Motor Impairments: A Position Paper. In *Proc. of the 15th Int. Conference on Human Computer Interaction (RoCHI '18)*. 99–106. https://dblp.org/rec/conf/rochi/GheranUV18

[33] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 623–635. https://doi.org/10.1145/3196709.3196741

[34] Elizabeth Goodman, Erik Stolterman, and Ron Wakkary. 2011. Understanding Interaction Design Practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1061–1070. https://doi.org/10.1145/1978942.1979100

[35] Tovi Grossman, Xiang Anthony Chen, and George Fitzmaurice. 2015. Typing on Glasses: Adapting Text Entry to Smart Eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. ACM, New York, NY, USA, 144–152. http://doi.acm.org/10.1145/2785830.2785867

[36] John Paulin Hansen, Alexandre Alapetite, Martin Thomsen, Zhongyu Wang, Katsumi Minakata, and Guangtao Zhang. 2018. Head and Gaze Control of a Telepresence Robot with an HMD. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications* (Warsaw, Poland) *(ETRA '18)*. ACM, New York, NY, USA, Article 82, 3 pages. https://doi.org/10.1145/3204493.3208330

[37] Alan Hashimoto. 2009. *Visual Design Fundamentals: A Digital Approach.* Course Technology, Cengage Learning, Boston.

[38] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy Norms and Preferences for Photos Posted Online. *ACM Transactions on Computer-Human Interaction* (March 2020). https://www.microsoft.com/en-us/research/publication/privacy-norms-and-preferences-for-photos-posted-online/

[39] MD. Rasel Islam, Doyoung Lee, Liza Suraiya Jahan, and Ian Oakley. 2018. GlassPass: Tapping Gestures to Unlock Smart Glasses. In *Proceedings of the 9th Augmented Human International Conference (AH '18)*. ACM, New York, NY, USA, Article 16, 8 pages. http://doi.acm.org/10.1145/3174910.3174936

[40] M.G. Kendall and B. Babington Smith. 1939. The Problem of m Rankings. *The Annals of Math. Statistics* 10, 3 (1939), 275–287. http://dx.doi.org/10.1214/aoms/1177732186

[41] Ron Kohavi and Roger Longbotham. 2017. Online Controlled Experiments and A/B Testing. In *Encyclopedia of Machine Learning and Data Mining*, Claude Sammut and Geoffrey I. Webb (Eds.). Springer US, Boston, MA, 922–929. https://doi.org/10.1007/978-1-4899-7687-1_891

[42] Quan Kong, Takuya Maekawa, Taiki Miyanishi, and Takayuki Suyama. 2016. Selecting Home Appliances with Smart Glass Based on Contextual Information. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 97–108. http://doi.acm.org/10.1145/2971648.2971651

[43] Lim Chee Koon. 2006. A Case Study of Icon-scenario Based Animated Menu's Concept Development. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services* (Helsinki, Finland) *(MobileHCI '06)*. ACM, New York, NY, USA, 177–180. http://doi.acm.org/10.1145/1152215.1152252

[44] Dong-Seok Lee and Wan Chul Yoon. 2004. Quantitative results assessing design issues of selection-supportive menus. *International Journal of Industrial Ergonomics* 33, 1 (2004), 41–52. https://doi.org/10.1016/j.ergon.2003.07.004

[45] L. Lee and P. Hui. 2018. Interaction Methods for Smart Glasses: A Survey. *IEEE Access* 6 (2018), 28712–28732. http://dx.doi.org/10.1109/ACCESS.2018.2831081

[46] Gitte Lindgaard, Janette Folkens, Catherine Pyper, Monique Frize, and Robin Walker. 2010. Contributions of Psychology to the Design of Diagnostic Decision Support Systems. In *Proceedings of HCIS 2010, the IFIP Human-Computer Interaction Symposium*. Springer, 15–25. https://doi.org/10.1007/978-3-642-15231-3_3

[47] MagicLeap. [n.d.]. MagicLeap One, Creator Edition. https://www.magicleap.com/magic-leap-one

[48] Meethu Malu and Leah Findlater. 2014. "OK Glass?" A Preliminary Exploration of Google Glass for Persons with Upper Body Motor Impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (Rochester, New York, USA) *(ASSETS '14)*. ACM, New York, NY, USA, 267–268. https://doi.org/10.1145/2661334.2661400

[49] Meethu Malu and Leah Findlater. 2015. Personalized, Wearable Control of a Head-Mounted Display for Users with Upper Body Motor Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 221–230. https://doi.org/10.1145/2702123.2702188

[50] Steve Mann. 2013. Vision 2.0. *IEEE Spectrum* 50, 3 (March 2013), 42–47. http://dx.doi.org/10.1109/MSPEC.2013.6471058

[51] Jérémie Melchior, Jean Vanderdonckt, and Peter Van Roy. 2011. A Model-based Approach for Distributed User Interfaces. In *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Pisa, Italy) *(EICS '11)*. ACM, New York, NY, USA, 11–20. http://doi.acm.org/10.1145/1996461.1996488

[52] Microsoft. [n.d.]. Microsoft HoloLens, The Leader in Mixed Reality Technology. https://www.microsoft.com/en-us/hololens

[53] Jurriaan D. Mulder. 2005. Menu Selection in Desktop Virtual Reality. In *Proceedings of the Central European Multimedia and Virtual Reality Conference (CEMVR '05)*, J. Sloup J. Zara (Ed.). Eurographics, Geneva, Switzerland, 1–7. https://pdfs.semanticscholar.org/724d/ba18240062929144d69f724fea00dbaaaaaf.pdf?_ga=2.248793991.1264527618.1584422185-1776159297.1548404692

[54] Mathieu Nancel, Stéphane Huot, and Michel Beaudouin-Lafon. 2009. Un espace de conception fondé sur une analyse morphologique des techniques de menus. In *Proceedings of the 21st International Conference on Association Francophone D'Interaction Homme-Machine* (Grenoble, France) *(IHM '09)*. ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/1629826.1629829

[55] K.E. Olson, M.A. O'Brien, W.A. Rogers, and N. Charness. 2011. Diffusion of Technology: Frequency of Use for Younger and Older Adults. *Ageing International* 36, 1 (2011), 123–145. https://dx.doi.org/10.1007/s12126-010-9077-9

[56] Irina Popovici and Radu-Daniel Vatavu. 2019. Understanding Users' Preferences for Augmented Reality Television. In *Proceedings of the 18th IEEE International Symposium on Mixed and Augmented Reality (ISMAR' 19)*. 269–278. https://doi.org/10.1109/ISMAR.2019.00024

[57] RedShed. 2018. The 8 Best Spy Glasses of 2018. https://redshed.co.uk/tech-reviews/best-spy-glasses/

[58] Christian Rohrer. 2014. When to Use Which User-Experience Research Methods. https://www.nngroup.com/articles/which-ux-research-methods/

[59] Peter H. Rossi and Steven L. Nock (Eds.). 1982. *Measuring Social Judgments. The Factorial Survey Approach.* SAGE Publications, Beverly Hills.

[60] Petruta Rusu, Maria-Doina Schipor, and Radu-Daniel Vatavu. 2019. A Lead-In Study on Well-Being, Visual Functioning, and Desires for Augmented Reality Assisted Vision for People with Visual Impairments. In *Proceedings of the 2019 E-Health and Bioengineering Conference (EHB '19)*. 1–4. http://dx.doi.org/10.1109/EHB47216.2019.8970074

[61] Krystian Samp. 2013. Designing Graphical Menus for Novices and Experts: Connecting Design Characteristics with Design Goals. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3159–3168. http://doi.acm.org/10.1145/2470654.2466432

[62] A. Santos, T. Zarraonandia, P. Díaz, and I. Aedo. 2017. A Comparative Study of Menus in Virtual Reality Environments. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 294–299. http://doi.acm.org/10.1145/3132272.3132277

[63] Marcos Serrano, Anne Roudaut, and Pourang Irani. 2017. Visual Composition of Graphical Elements on Non-Rectangular Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4405–4416. http://doi.acm.org/10.1145/3025453.3025677

[64] Snap. [n.d.]. Spectacles by Snapchat. https://www.spectacles.com/

[65] P.M. Sniderman and D.B. Grob. 1996. Innovations in experimental design in attitude surveys. *Annual Review of Sociology* 22 (1996), 377–399. https://doi.org/10.1146/annurev.soc.22.1.377

[66] Statista. 2017. Smart augmented reality glasses revenue worldwide from 2016 to 2022 (in million U.S. dollars). https://www.statista.com/statistics/610532/smart-ar-glasses-revenue-worldwide/

[67] Statista. 2017. Smart augmented reality glasses unit shipments worldwide from 2016 to 2022 (in 1,000s). https://www.statista.com/statistics/610496/smart-ar-glasses-shipments-worldwide/

[68] Yuya Takao, Michiya Yamamoto, and Tomio Watanabe. 2011. Development of Embodied Visual Effects Which Expand the Presentation Motion of Emphasis and Indication. In *Proceedings of the 14th International Conference on Human-computer Interaction (HCII'11)*. Springer-Verlag, Berlin, Heidelberg, 603–612. http://dl.acm.org/citation.cfm?id=2022466.2022537

[69] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3327–3336. http://doi.acm.org/10.1145/2702123.2702214

[70] Heather Turner and David Firth. 2012. Bradley-Terry Models in R: The BradleyTerry2 Package. *Journal of Statistical Software* 48, 9 (2012), 1–21. http://www.jstatsoft.org/v48/i09/

[71] Ovidiu-Ciprian Ungurean, Radu-Daniel Vatavu, Luis A. Leiva, and Daniel Martín-Albo. 2018. Predicting Stroke Gesture Input Performance for Users with Motor Impairments. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Barcelona, Spain) *(MobileHCI '18)*. ACM, New York, NY, USA, 23–30. https://doi.org/10.1145/3236112.3236116

[72] Jean Vanderdonckt, Mathieu Zen, and Radu-Daniel Vatavu. 2019. AB4Web: An On-Line A/B Tester for Comparing User Interface Design Alternatives. *Proc. ACM Hum.-Comput. Interact.* 3, EICS, Article 18 (June 2019), 28 pages. https://doi.org/10.1145/3331160

[73] Radu-Daniel Vatavu. 2012. User-Defined Gestures for Free-Hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video* (Berlin, Germany) *(EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626

[74] Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 224, 13 pages. https://doi.org/10.1145/3290605.3300454

[75] Radu-Daniel Vatavu, Bogdan-Florin Gheran, and Maria-Doina Schipor. 2018. The Impact of Low Vision on Touch-Gesture Articulation on Mobile Devices. *IEEE Pervasive Computing* 17, 1 (2018), 27–37. https://doi.org/10.1109/MPRV.2018.011591059

[76] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2019. Stroke-Gesture Input for People with Motor Impairments: Empirical Results & Research Roadmap. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300445

[77] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proc. of the 33rd ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 1325–1334. http://doi.acm.org/10.1145/2702123.2702223

[78] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support. In *Proceedings of the 2016 Conference on Human Factors in Computing Systems (CHI '16)*. 3390–3402. http://doi.acm.org/10.1145/2858036.2858228

[79] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap Gestures for TV: Insights from an Elicitation Study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '14)*. ACM, New York, NY, USA, 131–138. https://doi.org/10.1145/2602299.2602316

[80] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. ACM, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[81] Vuzix. [n.d.]. Vuzix Blade: Augmented Reality (AR) Smart Glasses for the Consumer. https://www.vuzix.com/products/blade-smart-glasses

[82] Florian Wahl, Martin Freund, and Oliver Amft. 2015. WISEglass: Multi-purpose Context-aware Smart Eyeglasses. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(ISWC '15)*. ACM, New York, NY, USA, 159–160. https://doi.org/10.1145/2802083.2808409

[83] Any Whitefield, Frank Wilson, and John Dowell. 1991. A framework for human factors evaluation. *Behaviour & Information Technology* 10, 1 (1991), 65–79. https://doi.org/10.1080/01449299108924272

[84] Tom Wilks. 2004. The Use of Vignettes in Qualitative Research into Social Work Values. *Qualitative Social Work* 3, 1 (2004), 78–87. https://doi.org/10.1177%2F1473325004041133

[85] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[86] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[87] Scott W. H. Young. 2014. Improving Library User Experience with A/B Testing: Principles and Process. *Journal of Library User Experience* 1, 1 (2014), 1–21. http://dx.doi.org/10.3998/weave.12535642.0001.101

[88] Mathieu Zen and Jean Vanderdonckt. 2014. Towards an evaluation of graphical user interfaces aesthetics based on metrics. In *Proceedings of the 8th IEEE International Conference on Research Challenges in Information Science (RCIS '14)*. 1–12. https://doi.org/10.1109/RCIS.2014.6861050

[89] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. ACM, New York, NY, USA, 239–249. http://doi.acm.org/10.1145/2700648.2809865