Why did you pick that? A study on smartwatch design qualities and people’s preferences

Raptis, Dimitrios; Papachristos, Eleftherios; Bruun, Anders Rysholt; Kjeldskov, Jesper

Published in:
Behaviour and Information Technology

DOI (link to publication from Publisher):
https://doi.org/10.1080/0144929X.2020.1836259

Publication date:
2020

Why Did You Pick That? A Study on Smartwatch Design Qualities and People's Preferences.

Dimitrios Raptis[*], Eleftherios Papachristos, Anders Bruun, and Jesper Kjeldskov

Human-Centered Computing Group, Department of Computer Science,

Aalborg University, Selma Lagerlöfs Vej 300, DK-9220 Aalborg Øst

{raptis, papachristos, bruun, jesper}@cs.aau.dk

* Corresponding author

Abstract: Several smartwatch vendors have over the past years entered the market with a large range of devices that differ in the design qualities they offer. In this paper, we focus on design qualities that are either perceived, descriptive or physical, and we investigate which of them may lead people to prefer particular smartwatches over others. To do so we conducted a laboratory study with 93 potential smartwatch users and a trained panel study with 8 participants. Results were analysed through a multivariate statistical technique called Preference Mapping. The advantage of this technique is that it can relate a large number of design qualities to users' preferences. Our findings show that participants can be divided into four groups with homogenous preferences, each emphasizing a unique combination of design qualities. For example, some groups emphasized device shape, while others prioritized expensiveness for their preferred smartwatch. We conclude with implications of our work on practice and research.

Keywords: Smartwatches, design qualities, perceived, descriptive, physical, coolness, preference mapping.

Word count: 12470

1 Introduction

One way to approach design is to view it as the process of embedding a set of design qualities into a product. Such design qualities can relate to materials, functionality, shape, color, etc. On the other hand, buying or not a product is the process of experiencing these designed qualities and forming a judgment about the product. During this process people do not usually try to thoroughly identify all design qualities, prioritize them, and rate them as researchers believed in the past. Instead, only the design qualities that are believed to be *relevant* for the specific context are taken onto consideration in forming a judgment (Van Schaik et al., 2012). For example, we know from existing research that the design quality of usability is favored in product choice situations, but only when people are asked about it (Diefenbach and Hassenzahl, 2009). Otherwise, different design qualities such as cost, beauty, functionality, brand and durability may be more relevant (Işıklar and Büyüközkan, 2007; Mack and Sharples, 2009; Sata 2013).

If a design quality that is considered relevant is missing, then people infer a value about it based on the ones they can identify (Van Schaik et al., 2012) by using pre-existing rules (also called heuristics) that are applied either consciously, or subconsciously. As previous research has shown, people can form first impressions extremely fast based on minimal information (Lindgaard et al., 2006). If a rule is going to be triggered or not in a specific context is related to people's past experiences, and the rule's effectiveness when applied in the past. For example, within user experience research a number of rules have been identified, such as "what is beautiful is good" (Dion et al., 1972), "what is beautiful is usable" (Hassenzahl and Monk, 2010; Tractinsky et al., 2000), and the reversed "it is usable therefore it is beautiful" (Tuch et al., 2012). The last two rules demonstrate that two design qualities may have a different relationship in different contexts.

Understanding which design qualities are relevant in specific contexts and how they trigger pre-existing rules to shape (often called drive) people's preferences is an important topic both for practitioners as well as Human-Computer Interaction (HCI) researchers. The relevance for practitioners is almost straightforward. Having such knowledge allows them to design better products and to better tailor them to the needs/preferences of specific target groups. For researchers, such knowledge is also important as it allows them to better understand how people experience digital products, which design qualities influence them and how, and how they shape their preferences.

A typical way to identify relevant influential design qualities that will trigger rules and shape people's preferences (and/or user experience) is through experimental studies. In such, a digital product's design quality (e.g. usability) is manipulated and its possible effect on user preference is documented. The process is then repeated with another experiment for another design quality for the same digital product in a specific context. Within HCI, there are numerous studies that reported on such experiments for a variety of digital products, such as mobile phones (Raptis et al., 2013), websites (Tuch et al., 2010), etc. Such experimental studies are very important as they contribute to our field's understanding of users' experiences with digital products by identifying cause and effect relationships among design qualities and/or preferences. At the same time though, they do pose a number of challenges. The first challenge is that the number of design qualities that can be simultaneously investigated through such experimental studies is relatively low, due to the complexity of experimental conditions. The second challenge is that it is not possible to holistically identify which design qualities are relevant in driving peoples' preferences in specific contexts.

In this paper, the products under investigation are smartwatches and we address those two challenges by providing two contributions. Firstly, we identify which smartwatch design qualities are considered as relevant in product choice situations through a laboratory study and a trained panel study. Secondly, we demonstrate how the identified relevant smartwatch design qualities trigger rules that shape people's preferences by utilizing an explorative statistical technique called Preference Mapping. Preference Mapping is a statistical technique that has been successfully used in a variety of fields (e.g. marketing), but its use in HCI is somehow limited.

Our paper is structured as follows. Firstly, we browse relevant literature for the case of smartwatches, and we provide details on how and why we conducted a laboratory study and a trained panel study. Secondly, we discuss how we analyzed the collected data through Preference Mapping and we present our findings. Finally, we conclude our work by discussing the implications of our findings both for practitioners as well as researchers. Furthermore, since Preference Mapping is a relatively unknown technique for the HCI field, and because we believe it could be relevant for investigating other digital products too, we discuss our reflections on using it.

2 Related Work

2.1 Perceived, Descriptive and Physical Design Qualities

Typically, when people are asked about their preferences towards a product, a variety of design qualities can be given as a reason. In general, design qualities can be classified based on their subjectivity and abstraction level into three types: a) subjective perceived qualities b) descriptive attributes, and c) objective physical qualities.

Subjective perceived qualities refer to people's attitude towards a product. For example, high perceived usability signifies that a person subjectively perceives a product as usable, even without interacting with it. Within HCI we have a lot of experience in dealing with such qualities, which we usually measure through established questionnaires. Examples of perceived qualities include hedonic and pragmatic qualities (Hasssenzahl et al., 2003: Van Schaik et al., 2012), classic and expressive aesthetics (Lavie and Tractinsky, 2004), or coolness (Bruun et al., 2016; Raptis et al., 2017; Sundar et al. 2014), etc. For reasons of simplicity for the rest of the paper we will refer to the subjective perceived design qualities as *perceived qualities*.

Descriptive attributes differ from perceived qualities in that they only describe a product's attribute without a value judgment about it. For example, when a dress is perceived as *formal*, people make a statement regarding its *style* without revealing any attitude towards it. For reasons of simplicity for the rest of the paper, we will refer to the subjective descriptive design attributes as *descriptive qualities*. Such qualities are usually very hard to be measured by novice users, since they do not have the experience to make purely descriptive judgments, and/or may not have the vocabulary or sensitivity to identify all the relevant descriptive qualities of a product. For example, while experiencing wine, novice wine consumers may not be able to identify viscosity or body as relevant in shaping their preferences. Within HCI, there have been studies which studied the effect of descriptive qualities on user's preference and/or perceived qualities (such as website symmetry, Tuch et al., 2010).

Finally, *physical qualities* differ from the previous two as they can be objectively and undisputedly measured. Such physical qualities relate to weight, size dimensions, etc. Within HCI we have studies that identified possible effects of physical qualities on users' preference and/or perceived qualities. For example, in Raptis et al. (2013) researchers studied the effect of the physical quality of screen size on users' preference and perceived usability for the case of mobile phones, or in Kim (2016) and Kim (2017) effects of screen shape and size have been identified for the case of smartwatches.

2.1 Smartwatch Design Qualities and Preferences

Research on how people realize specific design qualities and how these drive their experiences and preferences has been successfully conducted in the past for many products such as mobile phones (e.g. Chuang et al., 2001; Han et al., 2004; Ling et al., 2007; Yun et al., 2010). Besides mobile phones, an increasingly growing body of research is also investigating qualities that can influence user preferences, adoption and sustained use of wearable technology, such as smartglasses, smartwatches, and jewellery. Examples of such qualities include perceived value (Yang et al., 2016), body placement (Gemperle et al., 1998; Harrison et al., 2009), social comfortability (Dunne et al., 2014), functionality (Dunne 2010; Adapa et al., 2018), and fashion (Juhlin et al., 2016; Wang et al., 2017). Furthermore, within the domain of healthcare and fitness tracking it was identified that adoption of wearable devices can be influenced by perceived usefulness (Lunney et al., 2016; Chau et al., 2019), form, comfort of wearing the device, and battery life (Rantakari et al., 2016). In the same context, Canhoto and Arp (2017) found that functionality was the main driver for adoption of health and fitness wearables, while the drivers for sustained use were more related to data accuracy, portability and resilience.

Regardless of application area or wearable device type, aesthetic qualities are consistently identified as important (e.g. Juhlin et al., 2016; Dunne et al., 2014; Gemperle et al., 1998; Hsiao and Chen, 2018; Rantakari et al., 2016). Towards this end, Pateman et al. (2018) focused on individual differences regarding aesthetic preferences for wearable devices. In a three-part study, participants created their own wearable device based on their individual preferences. Those prototypes where then used by the participants for 5 days. The results underline the importance of aesthetic qualities in the use and continuous engagement with wearable devices and also indicate a need for customizability so that users can adapt the devices to their own preferences and needs.

Specifically for the case of smartwatches, most research work within HCI focuses on how people interact with them (Ashbrook et al., 2008; Chen et al., 2014; Cho et al., 2014; Gong et al., 2018, Huang et al., 2014; Klamka et al., 2020; Oakley et al., 2015; Ogata and Imai, 2015, Singh et al., 2018; Yeo et al., 2019; Wang and Grossman, 2020; Wong et al., 2020), their everyday usage (Pizza et al., 2006), and what motivates them to do so (Dehghani 2018). In relation to possible effects of smartwatch design qualities on preferences, we identified three relevant studies (Lyons 2015; Schirra and Bentley, 2015; Dehghani and Kim, 2019).

In detail, Schirra and Bentley (2015) conducted an interview study with five participants who owned and used a smartwatch for at least four months. The objective of that study was to identify what kind of applications mainly shape purchase decisions as well as issues related to everyday usage. Findings from the interviews revealed that several design considerations factored into choosing a smartwatch. For example, a smartwatch's aesthetic dissimilarity to a regular wristwatch, which was contrasted to "a horrible wrist communicator thing". Interviewees also referred to smartwatch designs in terms of *formal*, *casual*, or *sporty*. Physical size was also an important quality, particularly for the female interviewees, to whom it was challenging to find a smartwatch designed for them. For the current designs at the time (2015), it was stated: "they just look like a big, huge man's watch". One participant reported that she eventually found a smartwatch with a white and rose-colored metal band, which "best resembled a woman's fashion style watch". Towards that end, color was also considered important. Schirra and Bentley (2015) also found that all their interviewees had downloaded third-party apps such as Facebook, eBay and Swarm for their smartwatches, yet they were only used to a limited extent. Thus, some of the advanced features offered by a smartwatch over a regular wristwatch seemed unimportant.

Similarly, Lyons (2015) made a survey study with 50 respondents to understand factors going into buying and using a smartwatch. In order to do so, he asked respondents about their preferences in relation to the choice of digital “dumb watches” (Lyons 2015). The aim of the study was to transfer design considerations from the regular (but digital) wristwatch domain into the smartwatch domain. Like the study of Schirra and Bentley, results from Lyons’ survey showed that several design qualities drove the preference in selecting a particular watch. The design qualities that overlap with Schirra and Bentley’s findings, are *style* (in terms of formal/casual/sporty), *similarity to a regular wristwatch*, and *size*. Lyons also found that only a core set of features in the digital (dumb) watches were used. Features like timer, alarm, stopwatch and time zone facilities were not used often. Color was also one of the decisive design qualities for choosing a watch, along with sleek and simple styles, over flashy ones. Multiple respondents also owned more than one watch, since some were more suitable to use for sports.

Along the same direction, Dehghani and Kim (2019) conducted a study on how the appeal of a smartwatch may influence user’s experience and adoption. Smartwatch appeal was approached by three components, namely *design aesthetics*, *uniqueness* and *screen size.* Through a questionnaire study with 738 participants that assessed the effect of those three components on *purchase intention* and *use behavior*, they identified differences between current and potential users. For current users they identified a significant effect of design aesthetics and screen size on use behavior, and differences among male and female participants for the first case. For the potential users, design aesthetics and uniqueness had a significant effect on purchase intention.

Inspired by related work we chose to work with this research question: *“Which* *smartwatch design qualities are relevant and how do they shape people's preferences in product-choice situations?"*. To answer this research question, first we identified a set of smartwatches that we used in a laboratory study with the purpose to collect data on their physical qualities and on how people perceived them in general (Study 1). Then we conducted a trained panel study (Study 2) with the purpose to first identify descriptive qualities for any smartwatch that could be relevant in driving people's preferences, and then rate them for the same smartwatches as in Study 1. Finally, we analyzed all collected data through an explorative statistical technique called Preference Mapping.

3 STUDY 1: Preferences, Perceived and Physical Qualities

The purpose of the first study was to collect data on relevant physical and perceived smartwatch qualities as well as people's preferences. It was a laboratory study and in the following subsections we discuss in detail how we selected the smartwatches, the participants, and the measured design qualities as well as our results.

3.1 Selected Smartwatches

The first step in the process was to identify which smartwatches to include in our laboratory study. Since we knew beforehand that we would use Preference Mapping to analyze the collected data, we opted for the minimum number of smartwatches that would allow us to use this statistical technique. Thus, six smartwatches were chosen (Lavine et al., 1988). After carefully considering a variety of available smartwatches, a selection was made by keeping in mind that we had to include as much variety in design qualities as possible in order to make sure that we would be able to identify which ones are important in driving people's preferences. Figure 1 shows the included devices: Polar m600 (Device A), Motorola 360 2nd gen (Device B), Sony Smartwatch 3 (Device C), Zeblaze Blitz (Device D), NO.1 G4 (Device E), and NO.1 D6 (Device F).

A: Polar m600 **B: Motorola 360 2nd gen** **C: Sony Smartwatch 3**

D: Zeblaze Blitz **E: NO.1 G4** **F: NO.1 D6**

Figure 1. The 6 selected smartwatches.

Within this selection we opted for variability in smart watch design qualities. In detail, three smartwatches were chosen to be square (A, C, E) and three round (B, D, F), as it has been identified that the shape of a smartwatch face may influence how it is perceived (Kim 2016). Since size is important (Schirra and Bentley, 2015), three were relatively small (B, C, E) and three relatively big (A, D, F). Inspired by Lyons (2015) and Schirra and Bentley (2015) on style, two were selected to represent sport watches (A, D), two representing everyday watches (B, E) and two were selected as they appear formal (C, F). Three had a colorful idle screen face (B, D, E) and two a black-and-white one (A, C, F), as color can be an important design quality too (Schirra and Bentley, 2015). Two variations were also included in relation to their price. Thus, A, B, and C were considered expensive smartwatches (approx. 300, 200, and 180USD), while D, E, and F were considered relatively cheap (approx. 110, 60, and 45USD). All of them were Android smartwatches. A, B and C run on Android Wear, while D, E and F were based on Android 5.1. Finally, variation was also included in relation to the bracelets of which two were plastic (A, D), two leather (B, E), and two metallic (C, F).

Before conducting the laboratory study, three of the authors, informed by related work and by carefully examining the smartwatches, identified physical smartwatch qualities. All physical smartwatch qualities that could potentially have been relevant for driving people's preferences, have been included in the study. Table 1 presents the relevant physical qualities in detail along with their measurement units.

Physical Quality	**Description and measurement units**
Weight	The weight of the smartwatch in grams
Face Surface Area	The area of the smartwatch's face in cm^2
Profile Height	The thickness of the smartwatch in cm
Volume	The volume of the smartwatch in cm^3
Bracelet width	The width of the smartwatch bracelets at the locking mechanism in cm
Bracelet length	The length of smartwatch bracelets in cm

Table 1. Relevant physical smartwatch qualities and measurement units.

3.2 Participants

For our study, participants were recruited through social networks, through flyers around our university campus, and through direct contact by email. In total, 97 people agreed to participate in our study. Before the laboratory study, participants were asked to fill in a demographics questionnaire that contained information about their age, sex, prior experience with smartwatches, prior experience with wristwatches as well as their favorite mobile device brands. Four participants informed us that they already owned a smartwatch and they were excluded from the study. This allowed us to ensure that participants' preferences would not have been affected by everyday usage. Sixty-five participants informed us that they owned at least one regular wristwatch that they would wear every day. Furthermore, Apple, Samsung and One Plus were identified as the favorite mobile device brand for 77 participants. No one mentioned as their favorite brand the ones the 6 included smartwatches had, thus this allowed us to experimentally control for possible effects of brand (for effects of brand on perceived qualities see De Angeli et al., 2009; Rondeau 2005). In the end, we had 93 people participating in our laboratory study, 22 self-identified as female, and 71 self-identified as male, aged 19-31 (M=23.1, SD=2.73).

3.3 Measures

In order to collect data on relevant perceived qualities of smartwatches we chose to include the extended version of the Cool questionnaire (Raptis et al., 2017).

Our motivation for this choice was twofold. Firstly, we consider perceived coolness as an important quality in product-choice situations in general, and specifically for smartwatches as uniqueness may influence purchase intention (Dehghani and Kim, 2019). Secondly, this specific questionnaire has been created by taking into consideration established user experience questionnaires (Lavie and Tractinsky, 2004; Quinn and Tran, 2010; Sundar et al., 2014, Van Schaik et al., 2012, for details see Raptis et al., 2017). Table 2 presents the 6 perceived qualities that were measured using COOL Questionnaire.

Perceived Quality	Items	Scale
Desirability	4	1 to 7, strongly disagree - strongly agree
Rebelliousness	4	1 to 7, strongly disagree - strongly agree
Perceived usability	4	1 to 7, strongly disagree - strongly agree
Classic aesthetics	2	1 to 7, strongly disagree - strongly agree
Hedonic quality	3	1 to 7, strongly disagree - strongly agree
Overall coolness	3	1 to 7, strongly disagree - strongly agree

Table 2. Relevant perceived smartwatch qualities, number of items and measured scale.

3.4 Procedure

The participants entered along with a researcher in our usability laboratory in groups of six and the study was designed to last for about one hour. In order to mimic product-choice situations, the six smartwatches were placed on display stands on a table in the middle of the room and had the idle screens as depicted in Figure 1. Additionally, since brand can influence participants when rating perceived qualities (De Angeli et al., 2009; Rondeau 2005) all brand logos were masked.

The participants were not provided with specific tasks (action-mode, Hassenzahl and Ullrich, 2007) and were instructed to freely interact with all the smartwatches for about 10 minutes. This mimicked the real-world scenario of observing smartwatches inside a store, together with other customers. After making sure that all participants had a chance to feel and interact with all smartwatches, each participant was assigned to a specific laptop and independently provided ratings.

First, participants were asked to rank the smartwatches in relation to their preference. Preference data were collected on a linear, unmarked scale (from 0 to 100) with the verbal anchors “least preferred” and “most preferred” at the two ends. For this, we developed a web application in which participants could drag and drop the images of the smartwatches on the linear scale. Then, each participant individually rated all six smartwatches in a sequence through an online form that contained the Cool questionnaire (within-subjects design). The order of the questionnaire items was randomized, as was the smartwatches rating order.

At the end of each session, the researcher reset the smartwatches to the idle screen and placed them on the display stands to be ready for the next group of participants.

3.5 Study 1 Results: Physical Qualities, Perceived Qualities and Participants' Preferences

The physical qualities described in Table 1 were measured by three authors. Table 3 summarizes the final measurements.

Measured Physical Quality	A	B	C	D	E	F
Weight (gr)	64.00	54.00	134.00	69.00	102.00	54.00
Face Surface Area (cm^2)	12.24	13.85	22.00	19.63	14.51	17.02
Profile Height (cm)	1.40	1.20	1.10	1.40	1.40	1.20
Volume (cm^3)	17.14	16.62	24.20	27.48	20.31	20.42
Bracelet width (cm)	2.60	1.90	2.40	2.10	1.90	2.30
Bracelet length (cm)	6.50	5.00	6.10	5.20	5.50	5.30

Table 3. Measured physical qualities for the 6 smartwatches. A) Polar m600, B) Motorola 360 2nd gen, C) Sony Smartwatch 3, D) Zeblaze Blitz, E) NO.1 G4, and F) NO.1 D6.

Table 4 summarizes the results for the six measured perceived qualities for the six smartwatches in the form of averages from the 93 participants. Additionally, a reliability analysis was conducted for the measured perceived qualities. Cronbach α values were really high, ranging from .883 to .925.

Measured Perceived Quality	A	B	C	D	E	F
Desirability	2.88	**4.30**	3.16	2.98	3.00	2.45
Rebelliousness	3.22	**4.40**	3.26	4.28	3.55	3.34
Perceived usability	4.51	**5.47**	5.16	4.75	4.45	3.91
Classic aesthetics	4.30	**5.85**	4.96	3.55	4.21	3.67
Hedonic quality	3.31	**5.08**	3.67	4.23	3.47	2.89
Overall coolness	2.72	**4.52**	3.06	3.11	2.79	2.01

Table 4. Average scores for the 6 six measured perceived qualities for the 6 smartwatches (1-7 scale). Highest scores marked in bold. A) Polar m600, B) Motorola 360 2nd gen, C) Sony Smartwatch 3, D) Zeblaze Blitz, E) NO.1 G4, and F) NO.1 D6.

Participants rated the six smartwatches on an unmarked scale (from 0 to 100) from their least to their most preferred ones. Figure 2 presents the average preference score of 93 participants for each smartwatch. On average the most preferred smartwatch was B (73.58) and the least one was E (24.71). Furthermore, 47 participants selected smartwatch B as their most preferred, and 48 selected smartwatch E as their least preferred.

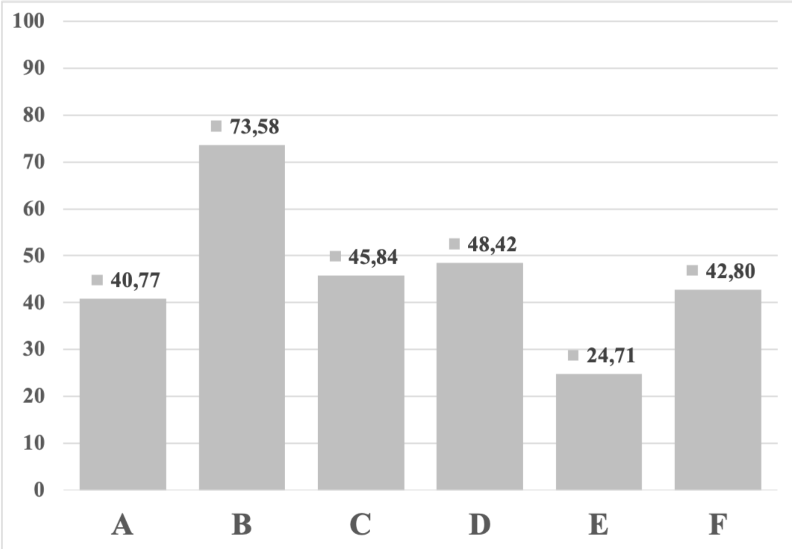

Figure 2. Average preference score of 93 participants for each smartwatch (scale 0-100). A) Polar m600, B) Motorola 360 2nd gen, C) Sony Smartwatch 3, D) Zeblaze Blitz, E) NO.1 G4, and F) NO.1 D6.

At the end of Study 1, we had ratings from 93 participants on perceived qualities for the six smartwatches, data about their preferences, and objective data on the six smartwatches' physical qualities. In general, smartwatch B (Motorola 360 2nd gen) scored the highest on all perceived qualities as well as participants' preferences. These results demonstrate that smartwatch B was the most preferred one, but they do not explain why this was the case. The following study sheds some light on the issue.

4 STUDY 2: Descriptive Smartwatch Qualities

The purpose of the second study was to extend related work by identifying additional relevant descriptive qualities for the case of smartwatches. Since descriptive qualities are difficult to realize for novice users, we opted for a trained panel study. Contrary to expert panel studies, in trained panels novice users are trained on how to identify and rate descriptive qualities, and typically after the initial training they provide comparable ratings to that of experts. Trained panels have been successfully used in a variety of disciplines such as sensory and food science (Lawless and Heymann, 2010), marketing (Urban and Hauser, 1993), and audio engineering (Mattila 2002).

4.1 Participants

Eight participants, 6 males and 2 females, aged 24-44 (M=32.75, SD=7.83) participated in the trained panel study. Since the purpose of the panel was to first identify and then rate any possible relevant smartwatch descriptive quality, it was important to include panelists with different backgrounds. Therefore, two of them were mechanical engineers, one was an electrical engineer, one was a visual designer, two were interaction designers, one was a usability expert and one was a techno-anthropologist. All of them were experts in technology-related fields, but none was an expert in smartwatches.

4.2 Procedure

Over the course of an eight-hour session, the panelists completed three phases. In the first phase, each panelist was provided with a small initial list of descriptive qualities, which were identified in the literature (Lyons 2015; Schirra and Bentley, 2015; Dehghani and Kim, 2019). Then, panelists were asked to remove, or add items to this list, based on how relevant they believed they were. Seven smartwatches (different than those used in Study 1) were also physically present in the room to act as inspiration for removing/adding descriptive qualities.

In the second phase, each of the identified descriptive qualities that each panelist created was repeatedly discussed and defined by all of them together. This process did not stop until all panelists reached a consensus on each identified descriptive quality's definition. At the end of the second part, a set of 36 descriptive qualities was produced, along with their semantic differential scales (e.g. *dull*/*bright*, on a 1-7 Likert scale).

Examples of the identified 36 descriptive qualities from the panel include *noisiness*, *bulkiness*, etc.

In the third phase each panelist individually rated the 6 smartwatches that were included in Study 1 over the 36 identified descriptive qualities, using an online form. The order they evaluated the 6 smartwatches was randomized, and so were the 36 identified descriptive qualities.

4.3 Study 2 Results: Identified Descriptive Qualities

The two main criteria for assessing a descriptive quality from a trained panel are discrimination ability and panelists' agreement. In general, a descriptive quality may not discriminate if participants are not able to perceive any differences among the products for this quality.

In our case, the discrimination ability of the 36 identified descriptive qualities was assessed through mixed-model ANOVAs with smartwatches as fixed factors and panelists as random ones. A descriptive quality was removed when a non-significant main effect was observed ($p>.05$).

Disagreement among panelists may occur if they understand a descriptive quality in a different way. Mixed-model ANOVA analysis can give a first indication about which qualities are problematic, but no single analysis method can give sufficient results on its own. For this reason, several univariate and multivariate inspection techniques (e.g. histograms, profile plots, eggshell plots (Hirst and Næs, 1994), Tucker-1 correlation plots (Dahl et al., 2008) have been applied to identify the qualities that created disagreement among panelists.

In the end, three descriptive qualities were removed due to low discrimination ability as they had a non-significant main effect. In addition, we removed another 11 qualities since multiple methods confirmed they caused disagreement among panelists. The remaining 22 descriptive qualities are presented in Table 5.

Descriptive Quality	Description and measurement scales
Shininess	How shiny the smartwatch is, without the face (glossy/matt)
Built Quality	The built quality of a smartwatch (fragile/robust)
Price	The price of the smartwatch (cheap/expensive)
Style	The style of the smartwatch (sports/formal)
Size	The size of the smartwatch (bulky/compact)
Complexity	How complex in terms of unnecessary design elements the smartwatch is (simple/complex)
Smartness	How much the smartwatch states it is smart (dumb watch/smart watch)
Waterproofness	How much water-resistant the smartwatch is (non-waterproof/waterproof)
Attention	How much attention the smartwatch attracts (modest/flashy)
Watch-Noisiness	How noisy the smartwatch is when shaken, without wearing it (not-noisy/noisy)
Felt-Temperature	How does the smartwatch feel when someone wears it (cold/warm)
Wristwatch-Prototypicality	How typical is the form of the smartwatch in relation to a wristwatch (non-typical/typical)
Shape	The shape of the smartwatch's face (round/square)
Color	How much color the smartwatch's idle screen has (colorless/colorful)
Brightness	How bright the smartwatch's idle screen is (dull/bright)
Resolution	How crisp the smartwatch's display is (grainy/crisp)
Swipe-Responsiveness	How responsive the smartwatch's display is, when swiping (non-responsive/ responsive)
Features	The number of features the smartwatch offers (few-features/many-features)
Bracelet-Traditionality	How traditional the bracelets' lock mechanism is (non-traditional/traditional)
Touch	How do the bracelets feel to the touch (harsh/soft)
Bendiness	How bendable the bracelet joints are (Rigid/flexible)
Button-Noisiness	How noisy the smartwatch's buttons are (when pressed) (not noisy/noisy)

Table 5. Identified descriptive smartwatch qualities and their semantic-differential scales.

5 Data analysis

In the following subsections we discuss the method we used to analyze the collected data from the two studies and our results.

5.1 Method

In order to analyze the collected data, we chose to use an explorative, visualization statistical technique called Preference Mapping (Carroll 1972; Meullenet et al., 2007). Preference mapping is very similar to another visualization technique called Multidimensional Scaling (MDS) (Carroll 1972; Schiffman, et al., 1981), which can be encountered in multiple HCI studies (e.g. Karapanos, et al., 2009; Wania, et al., 2006). Both techniques create a visual representation of various sources of collected data, but the key difference between MDS and Preference Mapping is that the former focuses on similarity and the latter on preference. Preference Mapping has been successfully used in many research domains such as marketing (e.g. Urban and Hauser, 1993; Van Kleef, et al., 2006) and food science (e.g. Helgesen et al., 1997), but its use in HCI is mostly in the domain of website aesthetics (e.g. Papachristos and Avouris, 2011; Papachristos and Avouris, 2013, Schenkman and Jönsson, 2000).

The outcome of Preference Mapping is a visual representation of a complex dataset in the form of a preference map. Such a map contains participants' preference ratings for multiple products (of the same type) as well as ratings of multiple product design qualities (perceived, descriptive and/or physical). This richness of projected data allows researchers and practitioners not only to have an overview of participant heterogeneity, but most importantly, to identify important drivers of preference by attempting to extract meaningful patterns out of these maps. The advantage of Preference mapping is that allows for the exploration of multiple design qualities at the same time by visually projecting correlations among these qualities and/or preferences. Since in our case we wanted to analyze data from multiple perceived, descriptive and physical smartwatch design qualities as well as participant's preferences, Preference Mapping was an ideal choice.

Two are the most prominent categories for Preference Mapping, namely external and internal (for an extensive review of these techniques see Carroll 1979; Meullenet et al., 2007; Van Kleef et al., 2006). In our data analysis, we used Internal Preference Mapping (IPM) since it is appropriate in cases in which prior knowledge about the importance of a specific quality is limited. The only prerequisite for conducting IPM is to have participant preference ratings for all the included products. This condition was fulfilled in our case.

In general, a preference map is created by using a data reduction technique (Meullenet et al., 2007; Naes and Risvik, 1996). For all preference maps we created, we conducted Principal Component Analysis (PCA) on a data matrix consisting of the six smartwatches in rows and participant preference ratings in columns. Before conducting PCA, as recommended by Greenhoff and MacFie (1994), we mean centered and standardized preference ratings to remedy for differences in use of scale by our participants.

5.2 Understanding Preferences for All Participants

5.2.1 Creating a Preference Map.

As a first step we created a preference map for the six smartwatches by including the preferences for all 93 participants. After conducting PCA, the first step was to assess how many principal components (or preference dimensions) to retain for analysis. In general, in Internal Preference Mapping researchers try to identify a small number of principal components (usually 2-3) that explain a large percentage in the variation of participant's preferences (Lawless and Heymann, 2010, p.442). In the case where two preference dimensions explain a large percentage of the total variance, it is very common to disregard the remaining dimensions for reasons of comprehensibility, if specific criteria are fulfilled (such as eigenvalues greater than 1, scree plot, interpretability: Lawless and Heymann, 2010, p.435). In our case, the first dimension accounted for 39.2% of the total variance, the second for 22.7%, while the third accounted for an additional 15.7%. After conducting a scree plot only the first two preference dimensions, which explained 61.9% of total variance, were included in the analysis. In Figure 3 the preference map of all 93 participants is presented. The two axes represent the two preference dimensions.

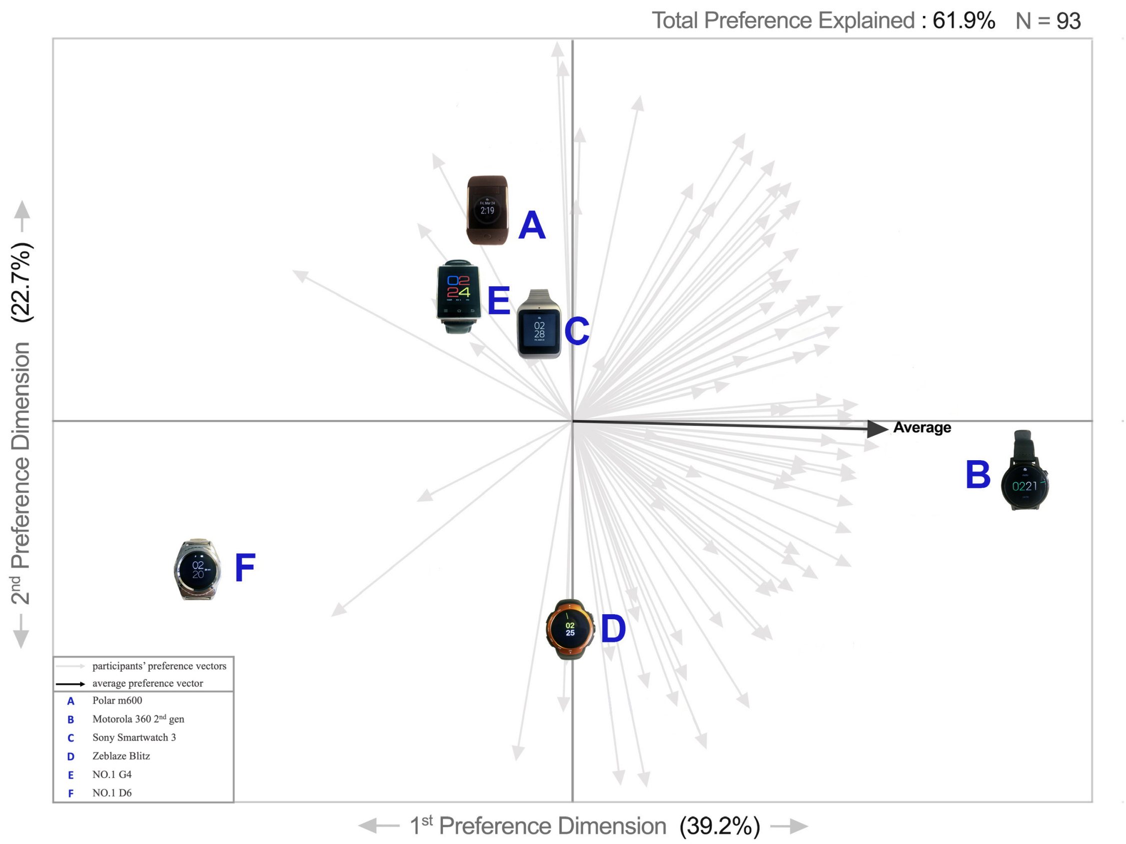

Figure 3. Preference map for the whole dataset. Grey arrows represent individual participant preference vectors while the black represents the Average.

5.2.2 Exploring preferences.

In Figure 3 the average preference of all 93 participants is denoted by a black vector with the caption 'Average', while each participant's preference is represented by a grey vector. The *length* of a vector indicates loading strength. Participants with low preference for any smartwatch are depicted close to the axis origin and have short vectors. The *direction* of the vector is also important. Preference vectors pointing to similar direction are positively correlated, while vectors pointing to opposite directions are negatively correlated.

From preference maps, as the one in Figure 3, we get knowledge about the products too, besides the participants. If we focus on the smartwatches, the proximity of devices A, C, and E is an indication that people's preferences for them were similar. On the other hand, devices B and F are far apart from each other because they received dissimilar preference ratings. Furthermore, devices A, C, D, and E have a little impact on the first preference dimension, as they are placed close to the vertical axis. The same is the case for device B and the second preference dimension.

If we observe Figure 3 closely, we will see that a great number of preference vectors are pointing to the general direction of device B. This means that participants who liked device B, at the same time disliked device F since it is positioned in the opposite side of their preference vectors. Furthermore, by examining the average preference vector, we can conclude that the average participants' preference vector (black vector) points towards device B, which makes sense, considering that the majority of participants' vectors point to the right side of the map. However, there is also a considerable number of preference vectors that diverge to the top (devices A, C, E) or to the bottom quadrant (device D) and very few that point to left of the map (device F). This indicates that different preference drivers existed among the participants.

5.2.3 Enriching the preference map with design qualities.

The challenge when trying to understand a preference map is to interpret the preference dimensions. This process is very subjective and usually the more data we project on a 2-dimensional map, the easier it is to interpret. In our case we enriched the previous map

(Figure 3) with the physical, perceived and descriptive qualities that we collected from the two studies.

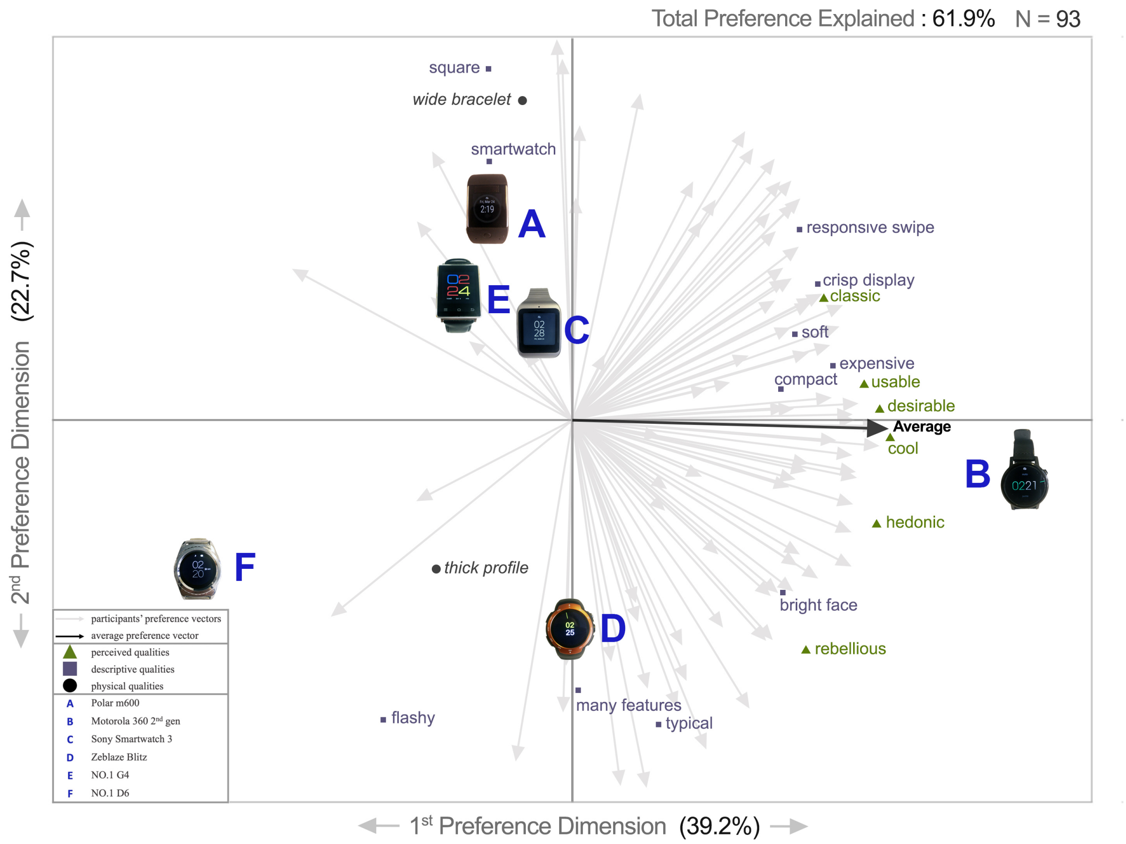

Figure 4. Preference map for the whole dataset this time with design qualities too. Grey arrows represent individual participant preference vectors while the black represents the Average. Green triangles denote perceived qualities, blue squares descriptive qualities, and black circles physical qualities.

Thus, a new preference map was created (Figure 4) by considering all the identified design qualities through a linear regression analysis (Greenhoff and MacFie, 1994), and by using their average score as dependent and smartwatch factor scores as independent variables. In general, only design qualities that have a significant relationship with preferences will appear on a map. Therefore, the design qualities that do not exist in Figure 4 did not have any significant relationship with preferences ($p>.05$). Each projected design quality is represented with a vector. If a design quality vector points to a specific direction (for example *compact*), then the opposing side defines the other end of the semantic differential scale (in this case *bulky*). We refer to Table 2, and Table 5 for the opposing ends of each semantic differential scale for each design quality.

5.2.4 Interpreting preference dimensions.

We can understand the relationship of two projected design qualities by investigating their angle. Small angle between them means that they point to a similar direction, and thus they are highly correlated. A 90-degree angle means that they are not correlated, while a 180-degree angle indicates negative correlation. For example, from Figure 4 we may deduct that the perceived qualities of *usable* and *desirable* are highly correlated, since the angle between the vectors is really small. Furthermore, the closer a smartwatch lies to the direction of a design quality vector, the more intensively it possesses that quality. For example, devices A, C and E are typical examples of a *smartwatch*, while the rest are not as they reside on the opposite side. Similarly, participants near a design quality vector like that design quality.

While trying to interpret the two preference dimensions from Figure 4, we can see that the six perceived qualities are all pointing to the right side of the map. However, the perceived qualities of *desirable*, *cool* and *usable* are all closer to the axes, and therefore are strongly correlated to each other as well as the first preference dimension. For explaining the first preference dimension, all devices except B and F provide little information, as they are positioned in the center of the horizontal axis and should be thought of as neutral. Thus, Device B was perceived to be more cool, more desirable and more usable than device F. The descriptive qualities that are more relevant for interpreting the first dimension are *compact* and *expensive*, since they are closer to it, and to a lesser extent *soft*, *crisp display*, *responsive swipe* and *bright face*. Similarly, we can assume that device B was perceived to be *cooler*, more *desirable* and more *usable* than device F, since it was more *compact* and *expensive* looking, while design F was rather bulky (opposite of *compact*) and appeared cheap (opposite of *expensive*). The second preference dimension is easier to explain since there are devices placed both at the top as well as the bottom part of the map. By observing Figure 4, it seems the second preference dimension is mostly defined by physical qualities, such as shape. Simple visual inspection reveals that smartwatches on the top are square and at the bottom round. In addition, devices on top have *wider bracelets* and look more than *smartwatches* rather than regular wristwatches. Devices on the bottom of the map are in the opposite end, as they look more as *typical* wristwatches, have narrower bracelets, are more *flashy*, and offer *more features*.

5.3 Understanding Preferences in Groups

Since we identified considerable diversity regarding the participants' preferences, it was not possible to identify a set of preference drivers that would be equally important to everyone. In such situations, it is fruitful to identify groups of participants with homogenous preferences and study them in isolation (McEwan et al., 1998). Our first step was to examine whether homogenous preference groups could be identified based on participant demographic data (e.g. sex, prior experience with wristwatches, etc.). However, this process did not reveal any significant results. Since visual recognition of distinct groups is not always easy as we may observe in Figure 4, we applied a hierarchical cluster analysis (Ward's method) on unstandardized preference data. Then we projected mean group preferences vectors in our Preference map to identify preference groups.

The end result was the identification of 4 distinct groups of participants with homogenous preferences. Separate preference maps have been produced for each of these groups by following the same process as before.

5.3.1 Group A: "The Idle-Screen Lovers"

Twenty-six participants were allocated to Group A, and two preference dimensions explained 81.6% of their variance. As it can be seen in Figure 5 all participants' preference vectors are pointing to the right side of the map, signifying an agreement among participants regarding the first preference dimension, which accounted for 63.9% of variance. Devices B, D and E are the most preferred, and A, F and C the least.

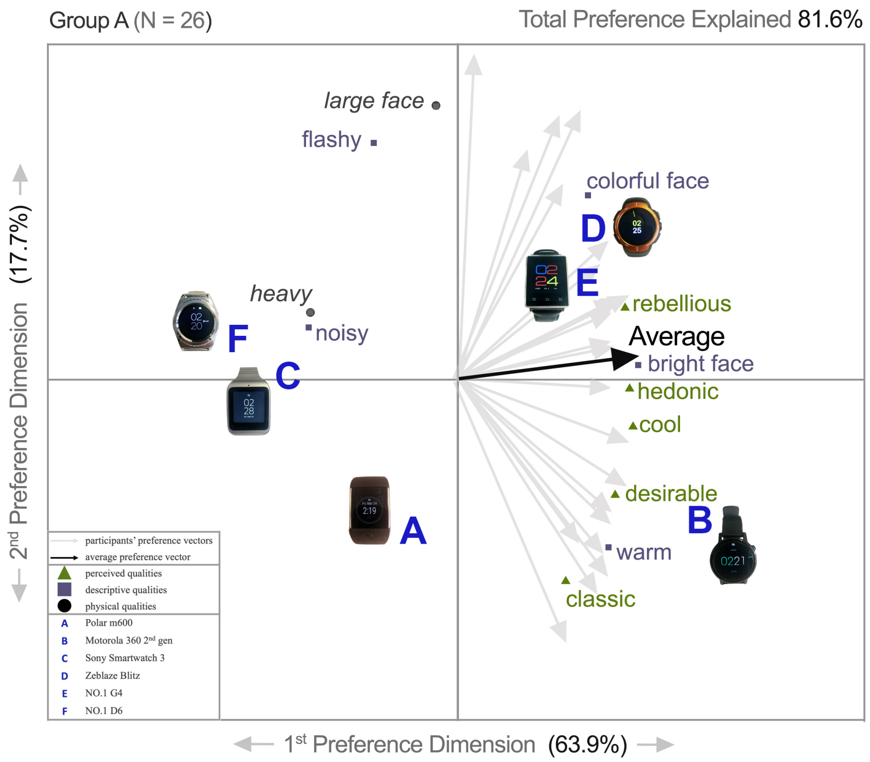

Figure 5. Preference Map of Group A: "The Idle-Screen Lovers".

The fact that the idle screen of the smartwatch had a *bright* and *colorful face* mainly drove the first preference dimension, and these are highly correlated with the perceived qualities of *rebellious*, *hedonic*, and *cool*. At the same time, the participants disliked the other smartwatches as they were perceived as *heavy* and *noisy*. *Large face area* and *flashiness* were important drivers for the second dimension, but since it accounts only for a small amount of variance (17.7%), it represents a less important preference driver. Based on these results, we call Group A the "Idle-Screen Lovers".

5.3.2 Group B: "The Premium Lovers"

Thirty-nine participants were allocated to Group B, and two preference dimensions explained 80.8% of variance. This was the largest identified group, and the one closer to the average preference of all participants. As it can be seen in Figure 6 all participants' preference vectors are pointing to the right side of the map, signifying an agreement among participants regarding the first preference dimension which accounted for 61.1% of variance. Devices B and C are the most preferred, while A, E, D and F the least.

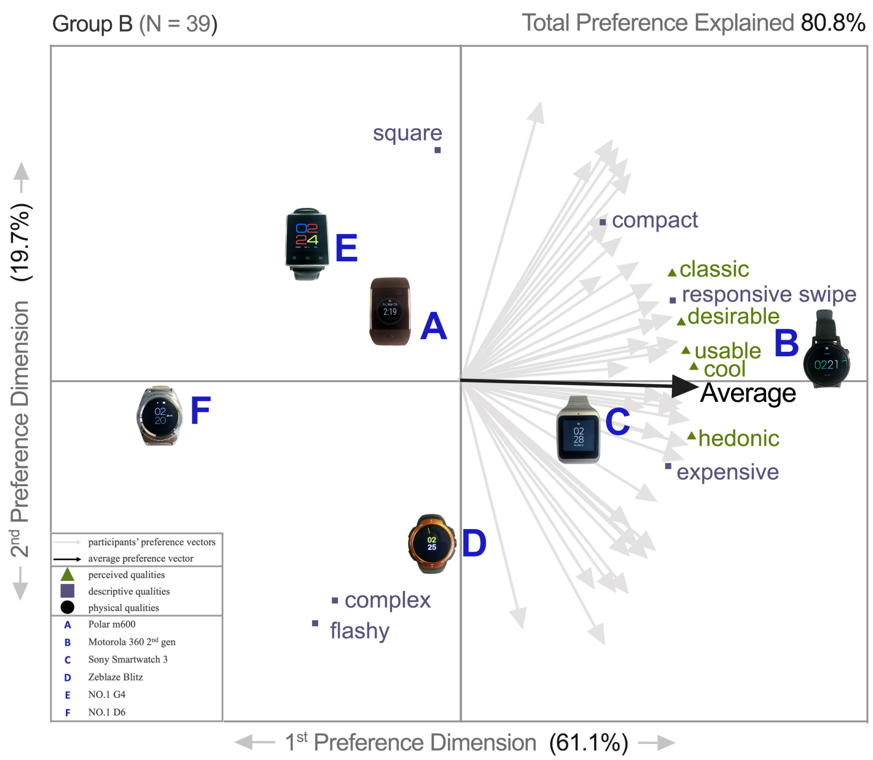

Figure 6. Preference Map of Group B: "The Premium Lovers".

Screen swipe responsiveness and *expensiveness* mainly drove the first preference dimension, and these are highly correlated with the perceived qualities of *usable*, *hedonic*, and *cool. Complexity* and *flashiness* were important drivers for the second dimension, along with *shape*, but since it accounts only for a small amount of variance (19.7%), it represents again a less important preference driver. Based on these, we call Group B the "Premium Lovers".

5.3.3 Group C: "The New-Form Lovers"

This is the smallest identified group, as only thirteen participants were allocated to it. Two preference dimensions explained 68.7% of the variance. As it can be seen in Figure 7 most participants' preference vectors are pointing to the right side of the map, signifying an agreement among participants regarding the first preference dimension which accounted for 44.1% of variance. Devices A, C and E are the most preferred, while B, D, and F the least.

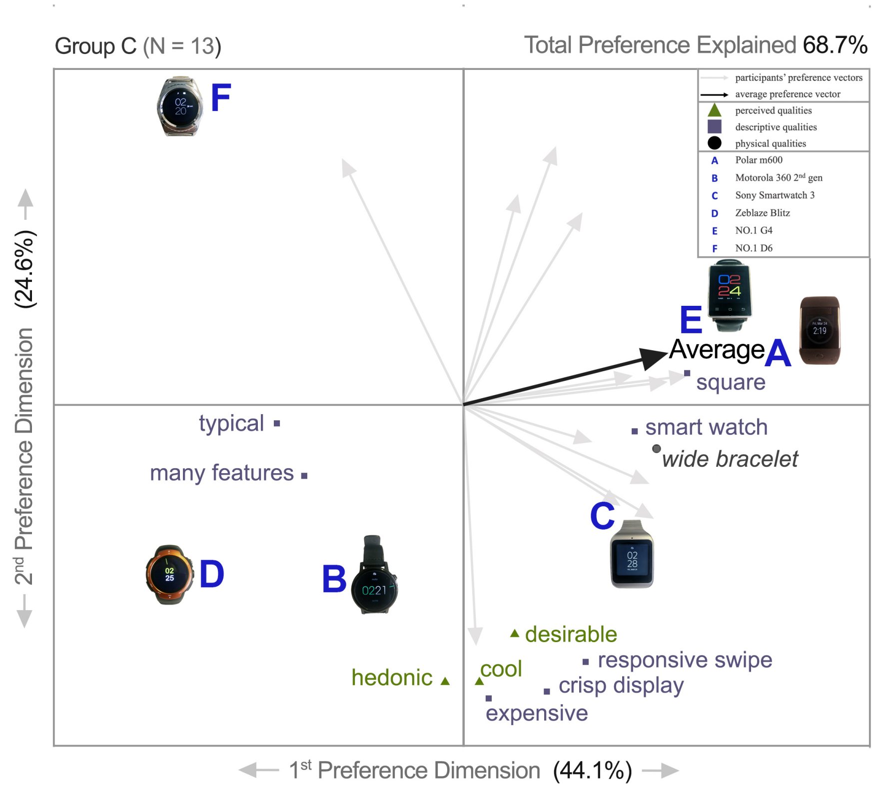

Figure 7. Preference Map of Group C: "The New-Form Lovers".

What is interesting in this group is that they identified the devices B, D, and C as *cool*, *hedonic* and *desirable* along with a number of descriptive qualities (e.g. *responsiveness*), but this did not influence their preference (their preference vectors point towards devices

E and A). Thus, their preferred devices were mainly selected because they were *square*, had *wide bracelets*, looked like a *smartwatch*, did not have *many features*, and their physical form opposed the one of a *typical* wristwatch. For this reason, we call Group C the “New-Form Lovers”.

5.3.4 Group D: “The Old-Form Lovers”

Fifteen participants were allocated to Group D, and two preference dimensions explained 82.4% of variance. As it can be seen in Figure 8 all participants’ preference vectors are pointing to the right side of the map, signifying an agreement among participants regarding the first preference dimension which accounted for 57.7% of variance. Devices B, F are the most preferred, while A, E, and C the least.

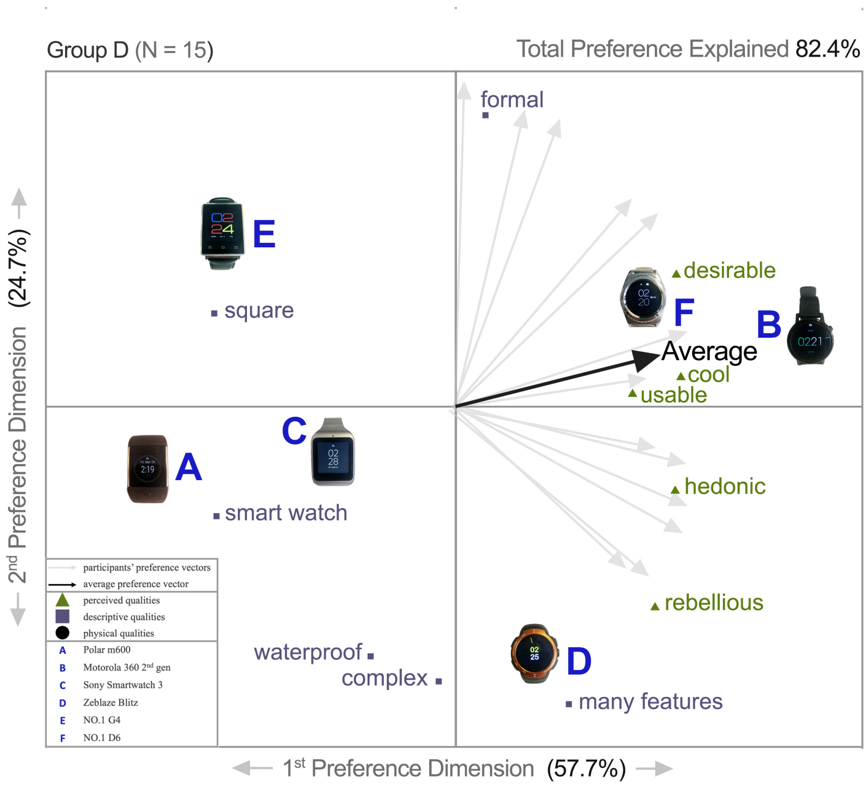

Figure 8. Preference Map of Group D: “The Old-Form Lovers”.

What is interesting in this group is that their preferences are driven by the opposite reasons of Group C. Thus, while Group C is driven by the typical form of smartwatches, participants of Group D preferred devices F and B as they were the exact opposites: they hide the fact they are *smart watch*es and they look like a typical round wristwatch. These descriptive qualities were also highly correlated with the perceived ones of *usable*, *desirable* and *cool*. Based on these, we call Group D "Old-Form Lovers".

6 Discussion

Our findings bring forward the two contributions of this paper: the identification of relevant smartwatch design qualities in product choice situations, and the demonstration on how they shape people's preferences. We will now discuss these two contributions in the form of implications for practice and research. Furthermore, since Preference Mapping is a relatively unknown statistical technique for HCI, we will also present our reflections on using it.

6.1 Implications for Practice

In relation to practice, our study provides four main implications. Firstly, the identified in the related work (Lyons 2015; Schirra and Bentley, 2015) design qualities of dissimilarity to a regular wristwatch, compactness, sleekness, idle-screen, low number of provided/used features and modesty (not flashy designs), were also present in our findings. Furthermore, we extended the work of Dehghani and Kim (2019) on the effect of smartwatch design aesthetics by breaking them down into specific, relevant design qualities (for example *responsive swipe*). This is important for practitioners as the identified relevant design qualities can be used as anchor points for their design efforts. For example, someone may choose to invest more on producing a smartwatch that has a responsive swipe as this may have an impact on some people's preferences.

Secondly, we linked the design qualities to 4 groups of people, where the preferences of each group were driven by different combination of design qualities. This shows that there is no single smartwatch design that will appeal to everybody. Extending the Dehghani and Kim (2019) study who identified that demographic variables may impact the use and purchase intention of a smart watch, our study shows that people's preferences can be grouped and explained by specific design qualities. We believe that this contribution can also guide practitioners in their design efforts. We suggest considering in some cases to design more narrow than broad; to focus more on the preferences of a particular group, than the average preference of everybody. For example, designers of new smartwatches can focus on the "Old-Form Lovers" group and produce alternative designs that have a more retro style by following the typical, round form of a wristwatch. Or focus on the "The Premium Lovers" group as it was the largest group identified and produce alternative designs that are perceived as expensive. We believe this way of thinking can be extended to other types of products, such as mobile phones, applications, websites, etc.

Furthermore, an important aspect of a design process is the evaluation of existing versions of a product in order to redesign it, and/or the study of existing, competitive products in order to understand the state of the art. For such activities, we suggest to practitioners to conduct studies similar to this one in order to collect insights for possible design directions. By combining design qualities (perceived, descriptive and physical) with preference ratings and by using preference maps designers can identify what works well and what not in an existing design, or even discover major preference drivers that no design is currently fulfilling. Additionally, if such maps are combined with qualitative data, such as interviews with selected participants (for example the ones that their preference deviated the most from the average), then we believe the impact of a design may be drastically increased.

Finally, our results showed that the Cool questionnaire (Raptis et al., 2017) can be considered a good measurement tool of perceived qualities for the case of smartwatches too as it provided good results, both due to its reliability as well as its highly correlation with participant's preferences. Although it is a relatively new instrument, such results indicate that practitioners can include it into their evaluation processes.

6.2 Implications for Research

An important part of user experience research is related to the identification of existing rules (or heuristics) that shape people's preferences in specific contexts (such as "what is beautiful is good", Dion et al., 1972). Our findings extend the related work, as we have indications of the existence of four rules (one per identified group) that potential smartwatch consumers apply in product choice situations.

Based on our results we have indications that Group A was driven by the rule "it has a nice idle screen; therefore, it is good", Group B by the rule "it feels premium; therefore, it is good", Group C by the rule "it feels like a typical smartwatch; therefore, it is good", and Group D by the opposite, "it feels like a regular wristwatch; therefore, it is good". We purposefully stated that we have indications that these rules exist since in order to have concrete results on the existence of such rules we need to conduct experiments and studies similar to Tractinsky et al. (2000), or Van Schaik et al. (2012).

Nevertheless, the identification of the 4 preference groups along with their corresponding design qualities and the possible rules people apply in product choice situations points to specific UX research directions. Firstly, it is almost certain that user heterogeneity exists to other products besides smartwatches. Thus, it is important to rethink how appropriate it is for user experience evaluations that we average the results from all our participants. Perhaps, in some cases we should consider having two-step evaluation process, where we first identify possible preference groups, and then we conduct separate evaluations per group.

Secondly, we believe that we need to also consider the existence of preference groups when we study relationships among specific design qualities. For example, in Raptis et al. (2013) researchers demonstrated that the screen size of a mobile phone has an effect on perceived usability. Wouldn't be interesting to know if this applies to all mobile phone users, or to only specific user groups? Thirdly, we believe it is important to conduct more research work in order to understand if and how the relationships among preferences and design qualities change over time, and preference maps may play an important role on this.

Finally, an important aspect of user experience research is the identification of measurable user experience qualities that can be useful to practitioners. If we observe the reported preference maps, we can see that in all cases except Group C, the perceived quality of coolness (Raptis et al., 2017) was highly correlated to the average preference vector (correlations between 0.7-0.9). This indicates that coolness may act as a mediator between specific design qualities and preferences. Of course, in order to have validated results there is a need to conduct experiments. Nevertheless, for user experience research, our results show that perceived coolness should be considered a relevant user experience quality that can be used both for better understanding what user experience is as well as for measuring it.

6.3 Reflections on Using Preference Mapping

Preference Mapping is not an appropriate statistical technique for revealing causal relationships among design qualities and preferences. Instead, it should be viewed as an ideal statistical technique for visually inspecting the relationships among such qualities and for exploring relatively unknown domains. Thus, it is a technique which is based on collecting quantitative data for the creation of the maps but understanding/explaining the maps is a highly subjective process. Having this in mind, we will present four ways in which Preference Mapping was useful for our research work, hoping that it can be useful to other HCI researchers and practitioners too.

Firstly, Preference Mapping proved extremely useful in providing *visual* representations of correlations among variables (only design qualities that have a significant relationship with preferences appear on a map). Visually inspecting a dataset was very insightful as it allowed us to explore the domain of smartwatches and have meaningful explanations relatively fast. We believe that this would not have been possible by following traditional statistical analysis techniques due to the large number of variables we included in our study. For this reason, we strongly recommend this technique to other researchers and practitioners, as visual inspection can sometimes be more informative than simply comparing numbers.

Secondly, Preference Mapping allowed us to combine data from different sources. In this research work we combined perceived, descriptive and physical design qualities with preference ratings. The technique allows for even more data sources to be superimposed on a map, allowing for better/richer explanations to emerge. This, however, may come at a cost in terms of complexity, since very often the more data points exist on a map, the more difficult it is to reach a meaningful explanation due to information clutter. Thus, we recommend to future users of this technique to create a variety of maps of varying levels of complexity until they reach a explanation they believe it is meaningful.

Thirdly, through Preference Mapping we managed to identify new interesting future research directions and we can now design a series of experiments that would allow us to identity possible causal relationships among specific design qualities and preferences. For example, one of those experiments could focus on the “Idle-Screen Lovers” group and study the cause and effect relationships between idle screen’s brightness/colorfulness, coolness and preference. Thus, approaching Preference Mapping as an idea generation technique for future research directions can be extremely useful for others too.

Finally, in order to properly do Preference Mapping there is a need for a large pool of participants (40-100 participants, Gagula Rutenbeck, 2006) and a considerable number of products (minimum 6, Lavine et al., 1988). Recruiting a large number of participants is never easy, but the most difficult task in this study was to select appropriate smartwatches, since we had to include a large variation in their design qualities. Our results show that we were successful in our selection as the smartwatches were spread across the maps. Nevertheless, this was a meticulous task as it took a lot of time to discuss possible candidate devices. For this reason, we strongly recommend to future users of the technique to allocate a considerable amount of time to their product selection processes.

7 Limitations

Our study has a number of limitations. At the time the study was carried out most smartwatch designs were predominately male-oriented both in their designs and their size. Thus, we faced the same situation as the Schirra and Bentley (2015) study. We tried to compensate by selecting a small, neutral smartwatch (Device B), but nevertheless in similar future studies, this has to be taken into consideration. Secondly, our sample was homogenous in relation to participant’s age, cultural background, and prior experience with smartwatches. Since it has been identified in prior research that there are differences between potential and actual users on how they perceive a smartwatch (Dehghani and Kim, 2019), our results are mainly relevant for young, potential smartwatch users.

8 Conclusions

In this paper, we make two contributions to the field of HCI. Firstly, we identified which smartwatch design qualities (perceived, descriptive and physical) are relevant in product choice situations. Secondly, we demonstrated how these design qualities shape people's preferences by analyzing the collected data using the statistical technique of Preference Mapping. Our findings showed that among young potential smartwatch users, there are four distinct groups in which different combinations of design qualities shape their preferences. Some of the participants are influenced by the idle screen of a smartwatch, most by how premium it is perceived, while many from its physical form (resemblance, or not to a typical round wristwatch).

These findings have implications both for practitioners as well as researchers, as our findings can be used both for the design of new smartwatches as well as for the evaluation of existing ones. Our implications for practice can be summarized by the identification of relevant smartwatch qualities in product choice situations. Our implications for research can be summarized by the identification of 4 possible rules that drive people's smartwatch preferences, which point to further research directions, as they should be validated through experimental studies. Furthermore, we demonstrated that Preference Mapping can be used to gain a deeper understanding on which design qualities may have an impact on users' preferences. We suggest that using Preference Mapping can be useful for studying other technologies too.

In the future, we aim to extend this work by researching in depth the possible relation between demographic characteristics and design qualities. Secondly, we want to follow up with experimental studies focusing on the 4 possible identified rules. Finally, we would like to examine how preferences, design qualities and coolness change over time.

Acknowledgements

We would like to thank all 93 people that participated in our laboratory study. Special thanks to the trained panel participants who dedicated so many hours for us.

References

Apurva Adapa, Fiona Fui-Hoon Nah, Richard H. Hall, Keng Siau, and Samuel N. Smith. 2018. Factors influencing the adoption of smart wearable devices. *International Journal of Human–Computer Interaction*, 34, 5, 399-409. https://doi.org/10.1080/10447318.2017.1357902

Dorit Alt and Meyran Boniel-Nissim. 2018. Using multidimensional scaling and PLS-SEM to assess the relationships between personality traits, problematic internet use, and fear of missing out. *Behaviour & Information Technology*, 37, 12 (July 2018), 1264-1276. https://doi.org/10.1080/0144929X.2018.1502353

Antonella De Angeli, Jan Hartmann, and Alistair Sutcliffe. 2009. The Effect of Brand on the Evaluation of Websites. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (INTERACT '09), 638-651. http://dx.doi.org/10.1007/978-3-642-03658-3_69

Daniel Ashbrook, Kent Lyons, and Thad Starner. 2008. An investigation into round touchscreen wristwatch interaction. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (MobileHCI '08), 311-314. http://dx.doi.org/10.1145/1409240.1409276

Anders Bruun, Dimitrios Raptis, Jesper Kjeldskov, and Mikael B. Skov. 2016. Measuring the Coolness of Interactive Products: The COOL Questionnaire. *Behaviour & Information Technology*, 35, 3 (February 2016), 233-249. https://doi.org/10.1080/0144929X.2015.1125527

Ana Isable Canhoto, and Sabrina Arp. 2017. Exploring the factors that support adoption and sustained use of health and fitness wearables. *Journal of Marketing Management*, 33, 1-2, 32-60. https://doi.org/10.1080/0267257X.2016.1234505

John D. Carroll. 1972. Individual differences and multidimensional scaling. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, R.N. Shepard, A.K. Romney, and S. Nerlove (eds). Seminar Press, 105–155.

Ka Yin Chau, Michael Huen Sum Lam, Mai Lai Cheung, Ejoe Kar Ho Tso, Stuart W. Flint, David R. Broom, Gary Tse, and Ka Yiu Lee. 2019. Smart technology for healthcare: Exploring the antecedents of adoption intention of healthcare wearable technology. Health Psychology Research, 7, 1. https://doi.org/10.4081/hpr.2019.8099

Xiang 'Anthony' Chen, Tovi Grossman, Daniel J. Wigdor, and George Fitzmaurice. 2014. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 159-168. https://doi.org/10.1145/2556288.2556955

Hyeonjoong Cho, Miso Kim, and Kyeongeun Seo. 2014. A text entry technique for wrist-worn watches with tiny touchscreens. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology* (UIST '14 Adjunct), 79-80. http://dx.doi.org/10.1145/2658779.2658785

Ming C. Chuang, Chien C. Chang, and Shang S. Hsu. 2001. Perceptual factors underlying user preferences toward product form of mobile phones. *International Journal of Industrial Ergonomics*, 27, 4 (April 2001), 247-258. https://doi.org/10.1016/S0169-8141(00)00054-8

Tobias Dahl, Oliver Tomic, Jens P. Wold, and Tormod Næs. 2008. Some new tools for visualising multi-way sensory data. *Food quality and preference*, 19, 1 (January 2008), 103-113. https://doi.org/10.1016/j.foodqual.2007.07.001

Milad Dehghani. 2018. Exploring the Motivational Factors on Continuous Usage Intention of Smartwatches among Actual Users. *Behaviour & Information Technology*, 37, 2 (January 2018), 145–158. https://doi.org/10.1080/0144929X.2018.1424246

Milad Dehghani and Ki Joon Kim. 2019. The effects of design, size, and uniqueness of smartwatches: perspectives from current versus potential users. *Behaviour & Information Technology*, 38, 11 (January 2019), 1143-1153, https://doi.org/10.1080/0144929X.2019.1571111

Sarah Diefenbach and Marc Hassenzahl. 2009. The "Beauty Dilemma": Beauty is valued but discounted in product choice. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09), 1419-1426. https://doi.org/10.1145/1518701.1518916

Karen Dion, Ellen Berscheid, and Elaine Walster. 1972. What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 3 (1972), 285-290. http://dx.doi.org/10.1037/h0033731

Lucy Dunne. 2010. Smart clothing in practice: Key design barriers to commercialization. *Fashion Practice*, 2, 1, 41-65. https://doi.org/10.2752/175693810X12640026716393

Lucy E. Dunne, Halley Profita, Clint Zeagler, James Clawson, Scott Gilliland, Ellen Yi Luen Do, and Jim Budd. 2014. The social comfort of wearable technology and gestural interaction. In *Proceedings of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC 2014), 4159-4162. https://doi.org/10.1109/EMBC.2014.6944540

Maximo Gagula Jr. and Sheri Rutenbeck. 2006. Sample size in Consumer Test and Descriptive Analysis. *Journal of Sensory Studies*, 21, 2 (April 2006), 129-145. https://doi.org/10.1111/j.1745-459X.2006.00055.x

Francine Gemperle, Chris Kasabach, John Stivoric, Malcolm Bauer, and Richard Martin. 1998. Design for wearability. In *Proceedings of Digest of papers. Second international symposium on wearable computers* (cat. No. 98EX215), 116-122. https://doi.org/10.1109/ISWC.1998.729537.

Keith Greenhoff and Hallidey J. MacFie. 1994. Preference mapping in practice. In *Measurement of food preferences*, H.J. MacFie and D.M.H. Thomson (eds). Blackie Academic and Professional, 137–166. https://doi.org/10.1007%2F978-1-4615-2171-6_6

Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang 'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang. 2018. WrisText: One-handed Text Entry on Smartwatch using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), Paper 181, 14 pages. https://doi.org/10.1145/3173574.3173755

Sung H. Han, Kwang J. Kim, Myung H. Yun, Sang W. Hong, and Jongseo Kim. 2004. Identifying mobile phone design features critical to user satisfaction. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 14, 1 (Winter 2004), 15-29. https://doi.org/10.1002/hfm.10051

Chris Harrison, Brian Y. Lim, Aubrey Shick, and Scott E. Hudson. 2009. Where to locate wearable displays? reaction time performance of visual alerts from tip to toe. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). 941–944. https://doi.org/10.1145/1518701.1518845

Mark Hassenzahl, Burmester, M., and Koller, F., 2003. Attrakdiff: Ein Frageboten zur Messung wahrgenommener hedonischer und pragmatischer Qualität. *Mensch & Computer 2003. Interaction in Bewegung*, 187-196.

Marc Hassenzahl and Daniel Ullrich. 2007. To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19, 4 (July 2007), 429-437. http://dx.doi.org/10.1016/j.intcom.2007.05.001

Marc Hassenzahl and Andrew Monk. 2010. The Inference of Perceived Usability From Beauty. *Human-Computer Interaction*, 25, 3 (2010), 235-260. http://dx.doi.org/10.1080/07370024.2010.500139

Hilde Helgesen, Ragnhild Solhei, and Tormod Næs. 1997. Consumer preference mapping of dry fermented lamb sausages. *Food Quality and Preference*, 8, 2 (March 1997), 97-109. https://doi.org/10.1016/S0950-3293(96)00037-7

David Hirst and Tormod Næs. 1994. A graphical technique for assessing differences among a set of rankings. *Journal of Chemometrics*, 8, 1 (February 1994), 81–93. https://doi.org/10.1002/cem.1180080108

Kuo-Lun Hsiao, and Chia-Chen Chen. 2018. What drives smartwatch purchase intention? Perspectives from hardware, software, design, and value. *Telematics and Informatics*, 35, 1, 103-113. https://doi.org/10.1016/j.tele.2017.10.002

Da-Yuan Huang, Ming-Chang Tsai, Ying-Chao Tung, Min-Lun Tsai, Yen-Ting Yeh, Liwei Chan, Yi-Ping Hung, and Mike Y. Chen. 2014. TouchSense: expanding touchscreen input vocabulary using different areas of users' finger pads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 189-192. https://doi.org/10.1145/2556288.2557258

Gülfem Işıklar and Gülçin Büyüközkan. 2007. Using a multi-criteria decision making approach to evaluate mobile phone alternatives. *Computer Standards & Interfaces* 29, 2 (February 2007), 265-274. http://dx.doi.org/10.1016/j.csi.2006.05.002

Oskar Juhlin, Yanqing Zhang, Jinyi Wang, and Anders Andersson. 2016. Fashionable Services for Wearables: Inventing and Investigating a New Design Path for Smart Watches. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (NordiCHI ’16), Article 49, 1–10. https://doi.org/10.1145/2971485.2971505

Evangelos Karapanos, Jean-Bernard Martens, and Marc Hassenzahl. 2009. Accounting for diversity in subjective judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 639-648. https://doi.org/10.1145/1518701.1518801

Konstantin Klamka, Tom Horak, and Raimund Dachselt. 2020. Watch+Strap: Extending Smartwatches with Interactive StrapDisplays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI ’20), 1–15. https://doi.org/10.1145/3313831.3376199

Ki Joon Kim. 2016. Round or Square? How Screen Shape Affects Utilitarian and Hedonic Motivations for Smartwatch Adoption. *Cyberpsychology, Behavior, and Social Networking,* 19, 12 (December 2016), 733–739. https://doi.org/10.1089/cyber.2016.0136

Ki Joon Kim. 2017. Shape and Size Matter for Smartwatches: Effects of Screen Shape, Screen Size, and Presentation Mode in Wearable Communication. *Journal of Computer-Mediated Communication*, 22, 3 (April 2017), 124-140. https://doi.org/10.1111/jcc4.12186

Talia Lavie and Noam Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 3 (March 2004), 269-298. http://dx.doi.org/10.1016/j.ijhcs.2003.09.002

Barry K. Lavine, Peter C. Jurs, and D.R. Henry. 1988. Chance classifications by non-parametric linear discriminant functions. *Journal of chemometrics*, 2, 1 (January 1988), 1-10. https://doi.org/10.1002/cem.1180020103

Harry T. Lawless and Hildegarde Heymann. 2010. *Sensory Evaluation of Food: Principles and Practices*. Springer-Verlag, New York.

Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and Judith M. Brown. 2006. Attention web designers: you have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25, 2 (2006), 115-126. http://dx.doi.org/10.1080/01449290500330448

Chen Ling, Wonil Hwang, and Gavriel Salvendy. 2007. A survey of what customers want in a cell phone design. *Behaviour & Information Technology*, 26, 2 (2007), 149-163. http://dx.doi.org/10.1080/01449290500128214

Kent Lyons. 2015. What can a dumb watch teach a smartwatch?: informing the design of smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (ISWC '15), 3-10. https://doi.org/10.1145/2802083.2802084

Abbey Lunney, Nicole R. Cunningham, and Matthew S. Eastin. 2016. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior*, 65, 114–120. https://doi.org/10.1016/j.chb.2016.08.007

Zoë Mack and Sarah Sharples. 2009. The importance of usability in product choice: A mobile phone case study. *Ergonomics*, 52, 12 (December 2009), 1514-1528. http://dx.doi.org/10.1080/00140130903197446

Ville-Veikko Mattila. 2002. Ideal Point Modelling of Speech Quality in Mobile Communications Based on Multidimensional Scaling (MDS). *Audio Engineering Society Convention*, Paper 5546 (April 2002).

Jean A. McEwan, P.J. Earthy, and C. Ducher. 1998. *Preference Mapping: A Review. Review No 6*. Chipping Campden.

Jean-François Meullenet, Rui Xiong, and Christopher J. Findlay. 2007. *Multivariate and Probabilistic Analyses of Sensory Science Problems*. Wiley-Blackwell.

Tormod Naes and Einar Risvik. 1996. *Multivariate analysis of data in sensory science*. Elsevier Science.

Ian Oakley, DoYoung Lee, Rasel M.D. Islam, and Augusto Esteves. 2015. Beats: Tapping Gestures for Smart Watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 1237-1246. http://doi.org/10.1145/2702123.2702226

Masa Ogata and Michita Imai. 2015. SkinWatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference* (AH '15), 21-24. http://dx.doi.org/10.1145/2735711.2735830

Matthew Pateman, Daniel Harrison, Paul Marshall, and Marta E. Cecchinato. 2018. The Role of Aesthetics and Design: Wearables in Situ. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18), Paper LBW518, 1–6. https://doi.org/10.1145/3170427.3188556

Eleftherios Papachristos and Nikolaos Avouris. 2011. The application of preference mapping in aesthetic website evaluation. In *Proceedings of IFIP Conference on Human-Computer Interaction* (INTERACT 2011), 616–619. http://doi.org/10.1007/978-3-642-23768-3_96

Eleftherios Papachristos and Nikolaos Avouris. 2013. The influence of website category on aesthetic preferences. In *Proceedings of IFIP Conference on Human-Computer Interaction* (INTERACT 2013), 445-452. http://doi.org/10.1007/978-3-642-40483-2_31

Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5456-5469. https://doi.org/10.1145/2858036.2858522

Jeffrey M. Quinn and Tuan Q. Tran. 2010. Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 353-362. https://doi.org/10.1145/1753326.1753380

Juho Rantakari, Virve Inget, Ashley Colley, and Jonna Häkkilä. 2016. Charting Design Preferences on Wellness Wearables. *In Proceedings of the 7th Augmented Human International Conference 2016* (AH '16), Article 28, 1–4. https://doi.org/10.1145/2875194.2875231

Dimitrios Raptis, Nikolaos Tselios, Jesper Kjeldskov, and Mikael B. Skov. 2013. Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services* (MobileHCI '13), 127-136. http://dx.doi.org/10.1145/2493190.249320

Dimitrios Raptis, Anders Bruun, Jesper Kjeldskov, and Mikael B. Skov. 2017. Converging Coolness and Investigating its Relation to User Experience. *Behaviour & Information Technology*, 36, 4 (2017), 333-350. http://dx.doi.org/10.1080/0144929X.2016.1232753

David B. Rondeau. 2005. For mobile applications, branding is experience. *Communications of the ACM*, 48, 7 (July 2005), 61-66. http://dx.doi.org/10.1145/1070838.1070867

Mesay Sata. 2013. Factors Affecting Consumer Buying Behavior of Mobile Phone Devices. *Mediterranean Journal of Social Sciences*, 4, 12 (October 2013), 103-112. https://doi.org/10.5901/mjss.2013.v4n12p103

Bo N. Schenkman and Fredrik U. Jönsson. 2000. Aesthetics and preferences of web pages. *Behaviour & Information Technology*, 19, 5 (2000), 367–377. http://doi.org/10.1080/014492900750000063

Susan S. Schiffman, M. Lance Reynolds, and Forrest W. Young. 1981. *Introduction to Multidimensional Scaling: Theory, Methods, and Applications*. Emerald Group Publishing.

Steven Schirra and Frank R. Bentley. 2015. "It's kind of like an extra screen for my phone": Understanding Everyday Uses of Consumer Smart Watches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15), 2151-2156. https://doi.org/10.1145/2702613.2732931

Gaganpreet Singh, William Delamare, and Pourang Irani. 2018. D-SWIME: A Design Space for Smartwatch Interaction Techniques Supporting Mobility and Encumbrance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), Paper 634, 13 pages. https://doi.org/10.1145/3173574.3174208

Shyam S. Sundar, Daniel J. Tamul, and Mu Wu. 2014. Capturing "cool": Measures for assessing coolness of technological products. *International Journal of Human-Computer Studies*, 72, 2 (February 2014), 169-180. https://doi.org/10.1016/j.ijhcs.2013.09.008

Noam Tractinsky, Adi S. Katz, and Dror Ikar. 2000. What is beautiful is usable. *Interacting with Computers,* 13, 2 (December 2000), 127-145. https://doi.org/10.1016/S0953-5438(00)00031-X

Alexandre N. Tuch, Javier A. Bargas-Avila, and Klaus Opwis. 2010. Symmetry and aesthetics in website design: It's a man's business. *Computers in Human Behavior*, 26, 6 (November 2010), 1831-1837. https://doi.org/10.1016/j.chb.2010.07.016

Alexandre N. Tuch, Sandra P. Roth, Kasper Hornbæk, Klaus Opwis, and Javier A. Bargas-Avila. 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior*, 28, 5 (September 2012), 1596-1607. https://doi.org/10.1016/j.chb.2012.03.024

Glen L. Urban and John R. Hauser. 1993. *Design and Marketing Of New Products* (2nd Edition). Pearson.

Ellen Van Kleef, Hans C.M. Van Trijp, and Pieternel Luning. 2006. Internal versus external preference analysis: An exploratory study on end-user evaluation. *Food Quality and Preference*, 17, 5 (July 2006), 387-399. https://doi.org/10.1016/j.foodqual.2005.05.001

Paul Van Schaik, Marc Hassenzahl, and Jonathan Ling. 2012. User-Experience from an Inference Perspective. *ACM Transactions on Human-Computer Interaction*, 19, 2, Article 11 (July 2012). https://doi.org/10.1145/2240156.2240159

Heetae Yang, Jieun Yu, Hangjung Zo, and Munkee Choi. 2016. User acceptance of wearable devices: An extended perspective of perceived value. *Telematics and Informatics*, *33,* 2, 256-269. https://doi.org/10.1016/j.tele.2015.08.007

Hui-Shyong Yeo, Juyoung Lee, Hyung-il Kim, Aakar Gupta, Andrea Bianchi, Daniel Vogel, Hideki Koike, Woontack Woo, and Aaron Quigley. 2019. WRIST: Watch-Ring Interaction and Sensing Technique for Wrist Gestures and Macro-Micro Pointing. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '19), Article 19, 1–15. https://doi-org.zorac.aub.aau.dk/10.1145/3338286.3340130

Myung H. Yun, Sung H. Han, Sang W. Hong, and Jongseo Kim. 2010. Incorporating user satisfaction into the look-and- feel of mobile phone design. *Ergonomics*, 46, 13-14 (May 2010), 1423-1440. https://doi.org/10.1080/00140130310001610919

Christine E. Wania, Michael E. Atwood, and Katherine W. McCain. 2006. How do design and evaluation interrelate in HCI research? In *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06), 90-98. http://dx.doi.org/10.1145/1142405.1142421

Jinyi Wang, Oskar Juhlin, and Nathan Hughes. 2017. Fashion Film as Design Fiction for Wearable Concepts. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA ’17), Article 49, 10 pages. https://doi.org/10.1145/3027063.3049782

Bryan Wang and Tovi Grossman. 2020. BlyncSync: Enabling Multimodal Smartwatch Gestures with Synchronous Touch and Blink. In *Proceedings of CHI '20: CHI Conference on Human Factors in Computing Systems* (CHI '20). https://doi.org/10.1145/3313831.3376132

Pui Chung Wong, Kening Zhu, Xing-Dong Yang, Hongbo Fu. 2020. Exploring Eyes-free Bezel-initiated Swipe on Round Smartwatches. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI ’20). https://doi.org/10.1145/3313831.3376393

Figure and Table List