Smart Wearables or Dumb Wearables? Understanding how Context Impacts the UX in Wrist Worn Interaction

Vivian Genaro Motti
Volgenau School of Engineering
George Mason University – Fairfax, VA, USA
vmotti@gmu.edu

Kelly Caine
School of Computing
Clemson University – Clemson, SC, USA
caine@clemson.edu

ABSTRACT
Recent advances in technology fostered the commercialization and usage of wearable devices. Among diverse form factors, wrist worn devices stand out. Benefitting from a conventional format and easy access, wrist worn devices, such as smart watches and fitness trackers, have been gaining popularity. While their continuous usage and close contact with the human body enable various applications, their limited computational resources summed with continuous changes in the context of use challenge designers in providing effective interactive solutions for end users. Seeking to understand how the context of use impacts the user experience and interaction with ten popular wrist worn wearables, in this study we analyzed the users' feedback: 545 users' comments were collected from Amazon, coded and aggregated. Based on the users' feedback, we identify 31 major problems that are currently faced in wrist worn interfaces. The analyses of the users' feedback led to a discussion about the causes and severity of those problems, and also to the definition of a set of design implications aimed at improving the user interaction with the next-generation wrist worn wearables.

CCS Concepts
• HCI Design • Interaction Paradigms • Contextual design • Ubiquitous and Mobile Computing.

Keywords
Wearables; Context-awareness; Interaction design; User experience.

1. INTRODUCTION
The versatility and continuous presence of wearable computers in users' lives extend the potential of wearables to support human activities in various domains. Wearables have become popular as activity trackers for human behavior (e.g. Fitbit), and as assistive technologies to support the user interaction (e.g. Thalmic Myo) and communication (e.g. Apple Watch). Recent improvements in technological components, including miniaturized sensors and more efficient power solutions also contributed to that.

Bringing technology closer to users' lives can provide several benefits to them, especially due to the continuous data collection and timely notifications. However, it also implies major drawbacks – the continuous usage and proximity to users can turn a wearable device in to a major inconvenience [3]; in particular if the user interfaces and interaction design of a wearable ignores the transient requirements imposed by dynamic changes in the users' contexts [7]. Wearable devices with constant notifications can be overwhelming for end users, distracting, disturbing, interrupting or even annoying them.

Technologies that are continuously worn, i.e. without or with only minor interruptions, must be *aware of* and *sensitive to* changes in the users' contexts. Being aware, properly interpreting and considering contextual information becomes mandatory for wearable interfaces to promote better user experiences, and consequently to foster the users' acceptance, adoption and sustained engagement with these technologies.

Despite being important in the wearable design, context-sensitive solutions are complex to implement. Contextual information is not only vast and heterogeneous, but is also subject to dynamic changes when users interact with a wearable on the go. To properly understand and consider contextual requirements during the design of a wearable, firstly, the users' habits and behaviors need to be well investigated and understood.

Seeking to identify, analyze and understand major contextual problems that users face in the interaction with existing wrist worn wearables, we collected critical reviews that users posted on Amazon (the largest e-commerce website in the United States). To analyze the users' perspectives, we selected reviews containing the user feedback about the interaction and experience with the ten most popular wrist worn devices commercially available. To focus on users' concerns related to the context of use, we coded the users' feedback and classified them according to the contextual factor (user, platform and environment) and the severity of the interaction problem (cosmetic, minor, major, catastrophic). We sought for design issues based on critical reviews: what users dislike in wrist worn interfaces, errors commonly faced, and interaction problems.

Our results indicate that a lack of contextual sensitivity is an important issue in the wearable interaction, users face problems with the incorrect identification of their activities, inaccuracy, poor screen readability, among others. Our results also suggest that a better understanding of the users' context, despite challenging, is key to improve the wearable design, especially for wrist worn wearables continuously used. In fact, most design issues users currently face are related to fitness trackers, as those are present daily in wearers' lives. Most design issues are also specific to output responses (e.g., graphic output responses unreadable under sunlight due to low contrast), and the inability to disambiguate contextual information (e.g., drumming or playing cards should not imply in more steps, while shopping, on the other hand implies in more steps even if users hold their hands still on the chart and no arm swings are detected). After classifying the problems identified based on their impact on the user interaction we propose a set of design implications.

SIGDOC '16, September 23 - 24, 2016, Silver Spring, MD, USA

ACM 978-1-4503-4495-1/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2987592.2987606

Table 1. An analysis of seven research studies on multimodal interaction designs for wrist worn wearables.

Study	*Goal*	*Modalities*	*Method*	*Experiment*
Gesture Wrist [15]	To propose a novel input method	Gesture-based for input (hand and forearm gestures) and tactile (vibration) for output	Technical implementation and validation	
Facet [6]	To propose a new input and output approach	Graphic User Interface for output, tactile (touch) and gesture (rotation) for input	Technical implementation and validation	
WatchIt [12]	To propose a new interaction technique	Tactile for input (gesture-based) on the wrist strap	Implementation, evaluation, and user study (pilot)	✔
Office Smart Watch [1]	To develop an smart watch application	Gesture-based (lock, knock, return) for input, and visual, audio and vibration for output	Exploration of wrist gestures for authentication and access, and technical implementation	
Edge Interaction [11]	To study touch as an input entry on the screen edge	Tactile for input (touch, gesture, strokes)	Exploration of the device edge for tactile input, through a prototype implementation and user study	✔
Watch Connect [4]	To define a toolkit for prototyping	Gesture-based and tactile interaction for input in smart watches (e.g., swipe, wave, knuckle)	Development of a toolkit for cross-device interaction with smart watches, design, prototyping, and test	
Shimmering Smart Watches [18]	To propose a Design Space for the user interaction	Graphic (LED lights) for output responses and notifications	Proposal and evaluation of alternative UIs (circle, point), with prototyping, and test	

Finally, we note that while some design issues are related to the exploratory stages of wrist worn technologies (due to a limited understanding of the users' contexts), others are linked to the lack of design guidelines, standards and best practices, and a poor user involvement during the development stages. We expect that a better understanding of the common interaction problems that users face with wrist worn wearables can primarily lead to improvements in the interaction design of the next-generation wearables, and consequently increase the adoption, acceptance and sustained engagement from end users.

2. WRIST WORN DEVICES

Wearable devices have significantly evolved in past decades. Miniaturized components, more efficient power solutions and flexible materials exemplify recent developments that boosted the creation and commercialization of novel form factors [14]. Although on-body interfaces have been available for decades, with the first studies dating from the early 2000 [15], recent advances in technology boosted the development of such devices, contributing to improve their interactive solutions, reduce costs and popularize them. Despite an increasing attention to wearables of all form factors, wrist worn devices stand out, not only because of their conventional format and similarity with a watch, but also due to their prompt access and suitability for diverse applications.

Although, wrist worn wearables are versatile concerning their potential applications, existing devices generally fall in to three categories: fitness trackers, smartwatches and armbands. Fitness trackers, such as Fitbit and Garmin, focus on sportive and athletic applications, counting steps, measuring overall physical activity and users' movements. Smartwatches, such as Apple Watch and Samsung Gear, serve as miniaturized computers, supporting and complementing smartphone functions, they provide message notifications, incoming calls, and alarms. Armbands, such as Thalmic Myo, focus on user interaction, detecting user gestures to support input commands. In this study, we focus on fitness trackers and smartwatches, not only because of the popularity of such devices but also because of their continuous usage and inherent opportunities for multimodal interaction.

2.1 Wrist Worn Interaction

The recent emergence and growing usage of wrist worn wearables led to various studies on the design of interfaces and interaction for such devices. These studies investigate multidimensional aspects of the wrist worn interaction, such as: input entry and output responses for multimodal solutions, graphic user interfaces, tactile, gesture-based and auditory approaches. Table 1 presents 7 studies focused on the user interaction with wrist worn wearables.

Rekimoto was one of the pioneers in exploring a wrist worn device for supporting the user interaction through gestures [15]. While a gesture-based approach was explored for input, vibration patterns were used for output responses. Lyons et al. (2012) also evaluated touch gestures for input entry in tactile screens [6], but focusing mainly on multidimensional GUIs that explore the wrist to extend the interactive surfaces of a bracelet.

Seeking to overcome visual occlusion and the fatty finger problem, Perrault et al. (2013) studied the tactile interaction using a watch strap [12]. In their experiment, auditory solutions were analyzed. Also focusing on gesture-based solutions, Bernaerts et al. (2014) studied knock, lock, and return gestures [1]. This study targets at office scenarios, in particular at granting access to physical rooms. Their application uses audio, vibration and graphics for output responses, and pre-defined gestures for input.

To assess the user interaction on the edge of a device, Oakley et al. (2014) built a prototype that employs touch sensors around a graphic display [11]. They explored the performance of end users in completing a set of tasks that required touch as input entry. Houben et al. (2015) also analyzed the user interaction around a device, providing a toolkit to facilitate the implementation of interactive solutions in smart watches and cross-device interaction [4]. The interaction space they defined includes: on the watch (via physical contact and touch), above it (through hand gestures) or in it (internal sensors collecting implicit or explicit gestures).

Table 2. Ten wrist worn wearables analyzed, three smart watches and seven fitness bands were selected based on their popularity: description of their features and total number of reviews / number of critical reviews, as published on Amazon.com by May 2016.

Apple Watch	***Samsung Galaxy Gear***	***Samsung Gear 2***	***Fitbit Charge***	***Garmin VivoFit***	***Misfit Shine***	***Samsung Gear Fit***	***Garmin Vivo Smart***	***Fitbit Blaze***	***Fitbit Alta***
									
Messages, call, browsing, fitness	Notifications (call, text, email), music control	Call notification, music, heart rate, step, calories	Steps, distance, calories, stairs, sleep, time, alarm	Steps, distance, calories, sleep, time, daily goals	Activity (walking, run, swim, cycle), sleep, time	Control call, messages, alarms, heart rate	Notification, time, progress, alerts, steps	Steps, distance, calories, stairs, heart rate, text, event, calls	Steps, distance, calories, workout, alarm, time
239 / 1,581	**329 / 1,285**	**369 / 1,506**	**3,948 / 11,648**	**1,804 / 5,396**	**1,893 / 4,689**	**713 / 2,007**	**872 / 1,913**	**334 / 1,807**	**357 / 1,445**
15%	**26%**	**25%**	**34%**	**33%**	**40%**	**36%**	**46%**	**18%**	**25%**

Xu et al. (2015) investigated minimalistic interfaces, analyzing the appropriateness of icons and lights for notifying end users (e.g., calendar event, call, message) [18]. Unlike the previous studies, they start with an analysis of commercial devices (e.g., Samsung Gear, Pebble and Nike Fuel band), regarding their input, output and computational capabilities.

Lowens et al. (2015) and Motti et al. (2015) proposed design recommendations and principles [5], [8], still such studies focused on a more empirical and theoretical approach and a cross-validation of the solutions proposed in real-world experiments is missing. Other related studies focus on the collaboration of users across devices, also using smart watches [2], [9].

The analysis of the related work shows that despite a deep investigation of multimodal interaction in wrist worn devices, for both input and output, most studies are exploratory, focusing on proposing and building wrist worn prototypes, rather than testing and evaluating those in real world scenarios. Table 1 highlights that out of the seven studies analyzed, only two (Edge-Interaction and WatchIt) involve also evaluations with end users. Still, those evaluations were limited to a controlled environment (lab studies).

To the best of our knowledge, only the work of Pizza et al. (2016) focused on the wrist worn interaction in the wild, by recording the user interactions in different situations [13]. Still, this work is limited to a small sample of participants, as only 12 persons participated. The authors also claim that they did not focus extensively on analyzing the watch interfaces and usability problems, however they remark issues with the lack of responsiveness of the device, task interruptions and navigation problems because the users considered the menus confusing.

2.2 Main Design Challenges for Wrist Worn Interfaces

The related work remarks numerous challenges involved in the design of interaction and interfaces for wrist worn wearables. Wrist worn wearables have a small screen size with limited interactive surfaces that restrict input and output options [6], [11]. The small hardware available in such devices also results in a weaker computational power and limited battery life [14]. The user-interaction with tactile GUIs is hindered with the visual occlusion caused by fatty fingers [12], and gestural input commands face issues with Midas gestures, i.e. unintended user gestures that trigger a system action [8]. Besides this, users rely on their familiarity with mobile phones to interact, which can confuse users because wrist worn wearables present fewer features and interaction options [13].

2.3 Lack of User Studies in the Wild

While the importance of considering context in the design of wearables is undeniable and long debated [7], [17], current work either lack user studies, or are limited to user tests in controlled environments usually conducted with a small sample of participants in a laboratory setting. By being executed in controlled environments, for instance as experiments in a laboratory with pre-defined interactive tasks and a small sample of participants, little is known about the user interaction in the wild, i.e. when the context largely varies and impacts significantly the user experience, hindering or even preventing the user's interaction.

3. METHOD

Because we were interested in assessing the contextual impacts in the wrist worn interaction according to a broad range of people who already had interest (named prospective users) or experience (named final users) in using such devices, and we wanted to gather a geographically and demographically diverse sample, we conducted an analysis of online comments posted by human users. To identify critical interaction problems, we first defined an online source to collect the data (Amazon, the largest retailer in the United States was chosen) and extracted the critical reviews published by users reporting their feedback about ten wrist worn devices (the most popular devices, based on the reviews' numbers, were selected). After selecting the ten most popular devices and the online source for data collection, we identified, analyzed and individually coded the users' concerns related to the context of use. A qualitative analysis (coding) was combined with a quantitative approach (frequencies of occurrences). Further details about the methodology chosen in this work are described in the following sections.

3.1 IRB Approval

To ensure the protection of human subjects, before we started the data collection and analysis, the Clemson University Institutional Review Board (IRB) approved this study as exempt.

3.2 Resources for Data Collection

The online source used for data collection was Amazon, the largest e-commerce website and online retail in the United States. Amazon has a diverse list of products available and provides retail services since 1995. The advantage of collecting the users' reviews through Amazon is the large sample of users, covering diverse opinions and perspectives, and the commenting feature of the website, that allows users to rank, review and provide their feedback about the devices bought and used. To minimize the potential bias of considering unrelated comments, a set of exclusion criteria was pre-defined (to remove irrelevant complaints, such as issues with the costumer service, post or warranty limitations of a given device).

The quality, as well as the nature, of the users' comments varies per individual user. While some users provide a more extensive, detailed and formal comments (e.g. highlighting the benefits and drawbacks of the device and technical aspects), other users post shorter comments, more informal and objective, not always serving for data analysis. Also, in the analysis of online contents, little is known about the demographic profiles of the users, except that they tend to be tech-savvy and use the Internet frequently (e.g. for shopping online and reviewing products). The comments posted are relatively recent, ranging from 2010 to 2016.

3.3 Device Selection

Based on their popularity and usage, we selected ten wrist worn devices that are commercially available nowadays. Then, we collected and analyzed the critical reviews published by users in the Amazon website. The ten devices chosen were selected according to the total number of reviews that users had posted online. Table 2 describes the ten devices, their features, percentage of critical reviews collected and the total number of reviews posted. The wrist worn devices selected can be broadly classified in two categories:

- **Smartwatches:** miniaturized computers that serve as a wearable accessory for smartphones, being often used for notifications of calls, messages, time and events. E.g.: Apple Watch, Samsung Galaxy Gear and Gear 2, Sony SWR50.
- **Fitness bands / Activity trackers:** bracelets dedicated to sense information about the users, such as: number of steps, sleep hours, and heart rate; they aim at raising users' awareness about their daily habits and lifestyles. E.g.: Fitbit (Charge / Alta / Blaze), Garmin (VivoFit / VivoSmart), Samsung Gear Fit, Misfit Shine, and Nike Fuel Band.

The number of critical reviews varied per device, ranging from 239 for the Apple Watch to 3,949 for the Fitbit Charge. In percentages, Garmin VivoFit had most critics (46%) and Apple watch the least (15%).

3.4 Data Collection

Each of the comments posted was read individually, and then if the contents were compliant with the inclusion criteria, i.e. relevant to understand the user interaction and experience, and problems related to interaction and contextual factors, they were manually extracted from the website and transferred to a spreadsheet. The exclusion criteria (e.g. comments complaining about the costumer services and warranty limitations of the device) were applied to prevent bias and to ensure that mainly relevant comments were considered. As exclusion criteria, the research team agreed to remove complaints about the costumer service and warranty limitations, as well as personal issues faced by users (e.g. the device was lost or stolen, or had delivery issues).

3.5 Data Analysis

To analyze the data in a qualitative approach, all contents transferred to the spreadsheet were read and individually coded. The primary codes corresponded to a pre-set of themes (contextual aspects: user, environment and platform), then a severity analysis was employed as secondary coding (more specialized). The contents associated to context (contextual information, context-awareness, and context-sensitivity) were selected from the data sample for further analysis (Table 3). We adopt Schilit et al. (1994) definition of context here [16], as any information that is (or can be) relevant to feed the system. Context here includes three dimensions: the user environment, his/her profile and technological resources of the device under analysis.

We focus our data analysis on the users' feedback about their interaction and excluded contents that were not related to the users' perspectives about a given device. Also, we focus our analysis primarily on contextual aspects of the end user interaction, due to the continuous usage of a wrist worn wearable, which implies in context changes and transient system requirements as well. For an application to be properly adapted to contextual changes, it needs first to be context-aware, i.e. identifying the context surrounding the user, and then context-sensitive, i.e. properly understanding the contextual information to use it in favor of the user interaction.

After the preliminary analysis and coding, the interaction problems identified in the user comments were classified according to four severity ratings [10]. Depending on the impact of the problem in the user interaction, we classified it in four categories. *Cosmetic* problems are those that do not prevent the user interaction but annoy the user. *Minor* problems bother the user but also do not prevent or hinder his/her interaction. *Major* problems are those that hinder significantly the user interaction, frustrating users, but still without preventing the main task completion. Finally, *catastrophic* problems are those that prevent the user from accomplishing his/her task and can put users in danger (e.g. when the user is driving and the device requires explicitly interaction and his/her attention for a long period to read all the UI contents). Table 4 describes the four severity rankings (inspired in Nielsen's classification from 1995) used to classify the problems identified as well as their definitions and examples.

4. RESULTS

After the data collection, each of the 545 comments was analyzed, associated with a code and aggregated based on their similarity. To propose respective design guidelines, the analysis of the comments focused on understanding what are:

(i) the most common problems users face in the wrist worn interaction, and
(ii) the major causes and severities of such problems.

According to the exclusion criteria, previously defined, we did not consider problems related to costumer-service, warranty, sales, defective devices, and personal issues (e.g. the device was lost or stolen). Instead, we focused on collecting and analyzing data on design and interaction problems that may hinder (or even prevent) the user experience and that could also be relevant to improve the user interaction and interfaces in next-generation devices. Overall, some of the most common aspects remarked in the users' reviews were: issues with the battery (low durability, charging and problems during setup), overall quality of the device (e.g., fragile, not sturdy, bulky), and trade-offs between costs and benefits (for instance concerning the usefulness of the device in users' lives – its purpose and features available versus the purchase price).

Table 3. Themes, Codes and Descriptors for contexts of use.

Theme	*Codes*	**Descriptors**
Context	**Environment**	Information related to the users' surroundings, including the levels of light, noise, and humidity (rain, sweat)
	User	Information related to the profile of the human user: preferences or eventual impairments (visual, cognitive, motor)
	Platform	Information related to the technological resources, including constraints, modalities for interaction, battery, etc.

While some problems were device-specific (e.g. the Fitbit Charge would often fall and get lost, and the Misfit Shine is difficult to place the battery correctly), other issues were recurrent across devices (e.g. the UI contents are difficult to read under sun light, and the personalization choices are limited). Some comments are also user-specific, concerning lack of customization features. Our analysis focuses especially on wrist worn interaction from a holistic approach, in other words, we are not focusing on the evaluation of a specific technology, device, brand, release or model, but instead, we focus primarily on problems that are commonly found by users in the wrist worn interaction (regardless of the device specificities).

4.1 Initial Coding – Contextual Typology

The qualitative analysis of the comments gathered aims at aggregating the data collected and identifying common aspects. To group the reviews, three contextual aspects have been considered, they are defined as follows.

The nature of the users' comments varies individually, according to their perspectives. Therefore, the interaction problems faced can also be analyzed from different perspectives, for instance: per device, modality of interaction, frequency of occurrence, severity levels, or major causes. In this work, we focus our primary analysis on contextual factors, i.e. whether the problems faced by the users were related to their environment, platform or profile. Still, those classes are not exclusive, and one problem may fall in to two or more categories. For instance, when the user cannot hear the auditory feedback of the device, the causes may involve: limitations of the sound amplification in the device itself, eventual hearing impairments of the user, or external noises of the user environment. For classification purposes, we focused on the major concern reported by the end user, i.e. his/her own rationale and justification when reporting the problem encountered.

Despite each device model having specific problems, some problems were present across devices, for instance inaccuracy in the step counting, limited battery life and missed notifications. In our analysis, we focus on interaction problems across devices.

In our analysis, most of the problems reported by users (n = 75%) were directly related to the platform, suggesting that the devices require improvements in their design to better accommodate the users' needs and contexts. Not only advances in state-of-the-art technological solutions are needed, but also more user involvement during the design phases is expected to achieve such improvements. The problems identified are presented as follows.

4.1.1 Environment

Problems related to the user environment concern how the situation and circumstances in which the user interaction takes place impacts his/her interaction. For example, if the user interaction occurs during a meeting, the output responses must be subtle, discrete and private. External factors of the users' context include difficulty to hear auditory feedback when the user is not located in a quiet environment. Despite the fact that a minority of the users' comments analyzed was coded as environmental problems (solely 8%), we note that the impacts of the problems reported in the user interaction tended to be higher (often preventing the user interaction and his/her task completion).

Readability issues due to specific light conditions: under sun light, for instance when users are running outdoors with a fitness tracker, and they need to use their hands to make shadow in the screen and be able to read the UI contents.

> *'...it is now impossible to see in the directly daylight at all. If you are running outside, there's no way to see the screen'* [P1, Fitbit Blaze user]

Or in low light conditions:

> '*The display: It is not backlit and is, again in MHO, difficult to read in low light*' [P2, Garmin VivoFit user]

Lack of Context-sensitiveness: when the circumstances and situations in which the user is located, turn the feature useless.

> *'I like the "move" as a motivator except if I am driving for a couple hours or sitting on a plane -with no option to move.'* [P3, Garmin Vivo Smart user]

Inadequate audio feedback (too silent or too loud): when noisy environments prevent the audio to be heard.

> *'The concept of using the watch for phone calls when your hands are busy is awesome, but the reality is you can't hear unless you are in a noise-free zone'* [P4, Samsung Galaxy Gear user]

Or the vibration of the silent-alarm is too loud, without any option available to change its settings:

> *'It says that it's silent. We all know that vibrating anything makes noise, but this thing is quite loud when it vibrates. Loud enough that it actually wakes my wife up'* [P5, Fitbit Charge user]

Adverse humidity levels: when users are swimming, diving, showering, sweating or in the rain), hindering the user interaction and experience with the (supposed-to-be water-proof) device, causing condensation under the screen.

> *'Not long after wearing it, I've noticed it constantly has condensation built up in the display. It's never been submerged in water, just sweat and rain.'* [P6, Fitbit Charge user]

Or ineffective sensing features under water:

> *'Heart rate monitor does NOT work in water - so if you're a swimmer like me, this is NOT the best tracker in the world.'* [P7, Garmin VivoFit user]

4.1.2 User

User-related problems refer to eventual disabilities (situational impairments) or users' impairments (e.g., low vision and small font size and icons), as well as specific preferences and requirements that users' may have regarding the customization and personalization of the device features (e.g., menu choices and functionality available). The user-related problems identified in the analysis of the comments selected corresponded to 11% of the reviews (n = 62).

Interaction issues: related to specific user impairments and disabilities, including motor or visual disabilities.

Due to motor impairments and user dexterity level:

'If you have any dexterity or mobility issues - don't get this. The clasp IS that difficult to close.' [P8, Fitbit Alta user]

Due to visual impairments:

'Font can be difficulty to read if someone needs reading glasses. Got band for my husband and he noted that the font size for the notification was a bit too difficult.' [P9, Garmin Vivo Smart user]

Steep learning curve: to understand how to use the device and get used to the input process frequently required.

The user adaptation to the device, its purpose and usefulness:

'for me it's just a pain to log every cup of water I drink when I already drink water all day long and know that I get plenty' [P10, Fitbit user]

Learning curve, remembrance and recall:

'I had to turn it on to "sleep" mode to monitor my sleep and I never remembered to do' [P11, Garmin VivoFit user];

Individual preferences: for features, formats and choices.

Lack of customization options:

'Notifications were nice but again pretty inflexible. It's an all-or-nothing situation, and often a single event on my calendar would result in multiple alerts on the device.' [P12, Garmin Vivo Smart user]

Lack of personalization features:

'Won't let me change my activity goal' [P13, Garmin Vivo Smart user]

Language issues (interoperability, translation or globalization):

'I also wish it would support other languages, when my mother texts me in Russian all I see on the Alta is squares.' [P14, Fitbit Alta user]

4.1.3 Platform

The problems associated with the platform often referred to the limitations in the technology or computational resources (e.g. battery drains quickly, internal memory is insufficient) or the fragility of the device (e.g. screen scratches easily, the color dies quickly). Problems related to the platform were the most frequent in our analysis, corresponding to 75% (n = 411) of the comments collected. In general, the lack of accuracy in the sensing algorithms was the major issue highlighted by users, especially regarding the (lack of) precision in measuring the users' steps, distance travelled, stairs climbed and heart rate levels.

Lack of accuracy: imprecision in measuring the users' data and properly interpreting their activities.

Inaccurate number of steps:

'The most disappointing function by far is the pedometer. It seems to track steps purely by arm swings, which means if you're pushing a shopping cart, mowing the lawn or even carrying something, you don't get credit for any walking. That's a little absurd for a modern lifestyle.' [P15, Samsung Gear Fit user]

Incorrect distance measurements:

'distance tends to be way off. Walking a trail I knew for a fact to be 4.5 miles ... my Vivosmart gave me a 2.3' [P16, Garmin Vivo Smart user]

Issues with the sleep tracking feature:

'the Charge decided I was sleeping while sitting at my desk working and then decided I was done sleeping at 3:00am when I actually woke up at 5:30am' [P17, Fitbit Charge user]

Problems with the heart rate measurement:

'The heart rate monitor wasn't very accurate at all. If I did three readings in a row the results would be random from 60-90 BPM on average.' [P18, Samsung Gear Fit user]

Interaction problems: for input entry in multiple modalities and with different approaches.

Menu and navigation issues:

'The sleep tracker has too many steps to use - there should be a 'getting up' button that is the first thing that pops up while you're in sleep mode. The last thing I want to do when I'm getting out of bed is swipe through three menu choices.' [P19, Samsung Gear Fit user]

Complex input entry solutions:

'you have to triple-tap the disc to tell it that you are starting to swim. And then, if you aren't sure if it has registered the triple tap (it flashes some tiny lights around the edges, which ... doesn't work well on the pool deck), and you triple-tap again, or forget to triple-tap, it WON'T COUNT THE LAPS. Instead, and worse, it counts your laps as steps, so it not only messes up your lap counts, it messes up your step count for the day, too.' [P20, Misfit Shine user]

Poor voice recognition:

'The only thing more embarrassing than talking to your watch ... is doing so and getting no response. Likewise, I'd respond to texts by dictating a response, just because I could, and I'm certain the receivers always knew it was dictated, as the punctuation was abominable, and often there was at least one misheard word. Oh well. Close enough. SEND IT! As if I had much choice. There was no way to change one word or fix punctuation on her tiny face' [P21, Apple Watch user]

Interaction problems for output: with multimodal responses, alerts or notifications.

Presentation and rendering issues in GUIs:

'...notifications were awkward to read. While you could change the orientation to be vertical or horizontal it seemed strange either way you looked at it. In vertical mode the big words would be hyphenated over several rows which made it even harder to read at times.' [P22, Samsung Gear Fit user]

Lack of alert, fails in calling the user attention:

'The text alert does work but without vibration or noise, it is useless. Of course you can argue that your smart phone has its own attention gathering displays, so Alta's text alert is superfluous (but is an advertised Fitbit Alta feature).' [P23, Fitbit Alta user]

Unperceivable responses due to weak vibration:

'the reason why I bought this gadget is for the call notification function. Although call notification works relatively well as far as lighting the LED display up each time I get a call, the band does not vibrate long nor hard enough to get your attention. Thus rendering

the call notification function useless.' [P24, Fitbit Charge user]

Or due to quick renderings:

'The display is very usable inside, although the messages disappear quickly. If you don't catch something the first time, it's gone. ... Since you don't have an opportunity to review or try again, I have now disabled that as a waste of battery.' [P25, Fitbit Alta user]

Problems to synchronize: or to connect the device with a smart phone.

Unstable connections:

'when opening Garmin's App ~ 70% of the time it wants to set up a new device or connect to one... even though the app is already connected/setup with my wrist band. With the APP showing "set up a new device" screen or saying device not connect is VERY misleading and this mostly is the result of all the returns I saw online.' [P26, Garmin Vivo Smart user]

Difficult to set the connection up:

'Getting it connected to Bluetooth was a challenge (and I'm technically astute). After quite a few tries, and reading in the user forum, I got it to connect. It turns out there are two Bluetooth connections.' [P27, Fitbit Blaze user]

Problems with battery: concerning the duration of power, compatibility, and frequency of charges.

Difficulties to insert the battery:

'the battery charge method. I find it very uncomfortable to push the module out of the frame, and even more uncomfortable forcing it back into the frame. Then the charge module is once again a unique device.' [P28, Fitbit Blaze user]

Limited battery life:

'having the phone connected to the watch really wears down the battery on the phone, which after 6 months has a pretty short life (less than 24 hours, even with minimal use) anyway.' [P29, Samsung Galaxy Gear user]

Lack of optimization and features to save the battery life:

'I was a little disappointed with the battery life. I had read that people were able to go 3 days before recharging. I'm not sure how they were able to do this. I have to recharge it almost every day.' [P30, Samsung Gear 2 user]

Task interruption for charging purposes:

'Lastly, you have to give up either sleep monitoring or a day of footsteps while it charges because the charger is on the back of the band so it must be removed to charge.' [P31, Samsung Gear Fit user]

4.2 Secondary Coding – Severity Levels

While some problems with the interaction simply hinder or delay the user interaction, other problems prevent the user interaction, the device functionality, and/or the task completion. To shed light in this concern, we analyzed the problems identified according to their severity, and respective impacts in the user interaction.

Table 4. Four severity ratings to classify the problems identified based on their impacts in the user interaction.

Severity	*Definition*	*Example*
Cosmetic	Distracts or annoys the user but still allows his/her interaction to continue smoothly.	When the feature is no longer needed (e.g. P3: move alerts when the user is flying).
Minor	Hinders the user interaction, by either delaying it or making it more difficult, annoying the end user.	When the settings for a feature are not available (e.g. P12: the notifications are inflexible).
Major	Hinders significantly the user task completing, disturbing and interrupting him/her in the main task, causing frustration or embarrassment.	The feature is not precise (e.g. P15: the step counter relies on arm swings).
Catastrophic	Prevents the user from accomplishing his/her task.	The functionality is interrupted when the device is charging (e.g. P31: the user gives up the sleep monitoring during recharge)

Each of the 31 interaction problems has been classified as: cosmetic, minor, major, or catastrophic (Table 4).

The nature of the problem has been used in the classification, as well as the frequency of the problem occurrence and the sentiment expressed by the users' in their comments. For instance, some users would classify a problem as a minor inconvenience while others would express disappointment, frustration, or discomfort as a consequence of the problem faced by them. The definitions of each class and examples are presented in the following sections.

4.2.1 Cosmetic Problems

Some of the interaction problems caused 'just' minor inconveniences for the users, being classified as cosmetic. By eliciting the users' problems we expect to lead to improvements in the design of future-generation devices, aiming especially at increasing the user satisfaction with a device rather than improving the users' effectiveness or efficiency in completing a task. Cosmetic problems were related to customization of the features available, i.e. allowing users to personalize the device features according to their own individual preferences, and to extend the original features of the device, for instance by having a larger number of applications available to install or to personalize the device features.

4.2.2 Minor Problems

Minor problems often cause a delay in the user interaction or make the users slightly uncomfortable, still without preventing them from completing their tasks or causing major interruptions. For instance, when users are unhappy about the standard duration of the battery life and they end up charging the device battery more often than desired, but still benefit from the device features.

4.2.3 Major Problems

Major problems are faced when the users feel disappointed, frustrated, or annoyed with the interaction; for example: most devices estimate the number of steps, what often leads to measurement errors.

Table 5. Classification of the 31 interaction problems according to their contextual factors and severity levels.

Severity	*Environment*	*User*	*Platform*
Cosmetic	3	-	-
Minor	-	12, 13	28, 29, 30
Major	2, 4, 5, 6	8, 9, 10	15, 16, 17, 19, 22, 26, 27
Catastrophic	1	11, 14	18, 20, 21, 23, 24, 25, 31

Because the calculations are often based on standard measures, without a precise calibration of the device and algorithm according to the users' profiles, it becomes almost impossible for the step counting to be always accurate. While an error rate does not prevent the user interaction with the device, it results in an inconvenience for most users.

4.2.4 Catastrophic Problems

Catastrophic problems have a strong impact on the user interaction, preventing users from accomplishing their main tasks. For instance, when the device vibrates but the intensity of such vibration is not strong enough to be perceived, the notification feature becomes useless. Also, when the sensing feature does not work depending on the users' context, the device is no longer needed, e.g. the heart rate sensor that does not work under water.

4.3 Quantitative Analysis

In a quantitative analysis we focus on the occurrences of the problems, first by contextual factor, and then by problem severity.

We calculate the percentages of the comments that were classified in each category and highlight the factors that lead to problems in the wrist worn interaction (due to environment, platform or user profile).

Problems associated with the platform were the most frequent in the user interaction, corresponding to 75% of the problems identified (n = 411). Interaction problems associated to users' and environmental concerns were the least frequent ones, corresponding to 11% and 8% respectively (n = 62 and n = 46). Comments that received no codes (due to the exclusion criteria) corresponded to 5% of the total (n = 26).

Concerning the classification based on the severity levels of the problems identified, most problems were classified as major (n = 14) and catastrophic (n = 10), while few problems were considered minor (n = 5) and just one problem was classified as cosmetic. Those frequencies did not vary significantly per contextual factor, as major and catastrophic problems were still the most frequent regardless of the associated factor.

5. DESIGN IMPLICATIONS

As a key factor to improve the interaction design, it is essential to identify and understand the users' contexts, as well as the specific circumstances in which the user interaction takes place. Considering the diverse situations of use, and their specific requirements and constraints in the interaction design, allows for better design decisions for the user interaction. Such consideration also enables the user interaction to occur more smoothly across different contexts of use when their transient requirements are accommodated. When the users' contexts and their respective constraints are unknown, the device must offer customization options to enable end users to adapt and personalize the interactive solutions, especially concerning the input entries, output responses and features available, according to the individual preferences and users' needs. Such adaptation is valuable across several design aspects, including for instance:

- the adaptation of auditory responses in volume, nature, and frequency;
- the adaptation of the vibratory responses for output in intensity, frequency and notification patterns;
- the choice of input entry commands concerning both the interaction sequence and interaction modality;
- the choice of menu items and features available according to specific users' preferences and needs;
- the change of the formats used to represent information; e.g. to represent the time, and according to the user language, measurement units and cultural aspects;
- the settings to modify the contrast level of the screen, enabling users to read the UI contents under sun light or in dark environments;
- the calibration of the sensing measurements according to the users' profiles and characteristics, especially when estimated data is used, e.g. for stairs climbed, descended or number of steps.

While such design implications may not cover exhaustively all the problems reported by users, they provide directions to tackle some of the most critical problems that users claim to face more frequently in the interaction with popular wrist worn wearables commercially available. By providing high-level implications, we focus on the users' perspectives about wrist worn technologies, their overall interaction and experiences with existing devices, and we also indicate directions to improve state-of-the-art solutions. At this point, we seek to propose implications that are technology-agnostic and have potential to suit different brands, models and releases of wrist worn devices.

6. DISCUSSION

Collecting online reviews for analysis of the users' feedback have many advantages – a large population sample, physically distributed, voluntary opinions, free format for expression, etc. However, it also comes with major drawbacks, such as: little is known about the demographic profile of the user population, their age and gender. We can assume that users who post online comments are often tech-savvy, English speakers, with regular Internet access and interest in technology. Still, it is not possible to assess the authenticity of the comments posted, i.e. they may be biased depending on the interests of the party (e.g. to focus on commercial purposes). In an attempt to overcome such issues, a large number of comments have been analyzed; still a cross-validation of the results with end users is necessary and is planned to be conducted as future work.

The users' comments analyzed have varied intensities, while some users tend to get very emotional about their devices and blame the companies (claiming to be 'addicted to' or 'dependent on' the technology), other users tend to blame themselves for interaction problems, these type of users also seems to adapt themselves and the use of the device to better fit their needs. For example, some users after noticing the device would fall, fixed the watch core with extra tape or a rubber band to prevent loss. Some users also report to *force* arm swing movements in order to get a more precise (or higher) number of steps.

There are several causes that lead to an interaction problem with a device, as problems reported range from synchronization, and accuracy, to skin rash and durability. We note that not all problems are necessarily caused by bad design or lack of user involvement in the development process, some of the problems are actually related to the limitations in technology, while others are a result of inaccurate algorithms and problems specific to the application design. The lack of (or poor) understanding of the users' contexts can be a major cause of those problems, as well as ambiguities inherent to the activity performed or even the position of the sensor on the human body.

In this study, we identify major issues that users face while interacting with a wrist worn wearable, regardless of brand, release or model, but especially concerning the contexts in which the interaction takes place and the impacts that contextual factors have on it. We note tough that not all problems identified can be easily solved by design improvements; some problems still require major advances in technology to be resolved.

7. CONCLUSION

Wearable technologies, and wrist worn devices in particular, provide a large range of applications that can be relevant for end users in different domains. Despite such a large potential and increasing popularity, wrist worn wearables are still an emergent technology with several open challenges to properly design their interfaces and interactive solutions. This occurs for numerous reasons, including the fact that wrist worn wearables are in close contact with the user and continuously used, besides having limited computational resources and limited design guidance too.

To shed light in how the context impacts the user experience with wrist worn wearables – smart watches and fitness bands in particular, and to identify the major problems that users currently face, in this work we analyzed 545 comments posted online by users in Amazon concerning the ten most popular smart watches and fitness bands commercially available in 2016. Our results indicate 31 interaction problems related to the users' contexts, mainly due to the platform, computational and technological issues. Our analysis also suggests that most problems are classified as major or catastrophic, leading to both the users' frustration and task interruptions. With the analysis of the 31 problems identified, we propose respective design implications. We expect that such implications can aid to improve next-generation solutions for wrist worn wearables in terms of user interaction and experiences with multimodal interfaces.

As future work, there are several venues foreseen to build upon the analysis of users' reviews. For instance, concerning the methodology, we noted that the analysis of the users' comments provided us valuable insights about the semantic, lexicons and terminology that users employ to express their opinions on interaction problems. Such knowledge can aid to perform other reviews, as well as to optimize the evaluation process. Moreover, the design implications can be relevant to feed the development process as guidelines, still a validation study with designers and end users is necessary to assess their actual implications and validity concerning the usability levels of wrist worn devices.

8. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1314342. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

9. REFERENCES

[1] Bernaerts, Y., Druwé, M., Steensels, S., Vermeulen, J., and Schöning, J. 2014. The office smartwatch: development and design of a smartwatch app to digitally augment interactions in an office environment. In *Proceedings of the 2014 companion publication on Designing interactive systems* (DIS Companion '14). ACM, New York, NY, USA, 41-44. DOI=10.1145/2598784.2602777 http://doi.acm.org/10.1145/2598784.2602777

[2] Chen, X., and Grossman, T. 2014. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the 32nd SIGCHI Conference on Human Factors in Computing Systems* (CHI 2014), 159–168. DOI=http://doi.org/10.1145/2556288.2556955

[3] Dunne, L. E., Gioberto, G., and Koo, H. 2011. A Method of Measuring Garment Movement for Wearable Sensing. In: *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers* (ISWC '11). IEEE Computer Society, Washington, DC, USA, 11-14. DOI=10.1109/ISWC.2011.19

[4] Houben, S., Brudy, F., and Marquardt, N. 2015 Challenges in Watch-Centric Cross-Device Applications. Mobile Co-Located Interactions Workshop. In *Proceedings of the 33rd SIGCHI Conference on Human Factors in Computing Systems* (CHI 2015)

[5] Lowens, B., Motti, V. G., and Caine, K. 2015. Design Recommendations to Improve the User Interaction with Wrist Worn Devices. *WristSense: Workshop on Sensing Systems and Applications Using Wrist Worn Smart Devices*. St. Louis, MS. 562-567 DOI=10.1109/PERCOMW.2015.7134099

[6] Lyons, K., Nguyen, D., Ashbrook, D., and White, S. 2012. Facet: A Multi-Segment Wrist Worn System. In *Proceedings of the 2012 Annual ACM Symposium on User Interface Software and Technology* (UIST '12), 123–130.

[7] Lyons, K., and Profita, H., 2014. The Multiple Dispositions of On-Body and Wearable Devices, *Pervasive Computing*, IEEE, vol. 13, no. 4, pp. 24, 31, Oct.-Dec. 2014 DOI= 10.1109/MPRV.2014.79

[8] Motti, V. G., and Caine, K. 2015. Micro Interactions and Multi dimensional Graphical User Interfaces in the Design of Wrist-Worn Wearables. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* September 2015 vol. 59 no. 1 1712-1716.

[9] Nebeling, M., To, A., Guo, A., Freitas, A. A. De, Teevan, J., Dow, S. P., and Bigham, J. P. 2016. WearWrite : Crowd-Assisted Writing from Smartwatches. In *Proceedings of the 34th SIGCHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 3834-3846. DOI=http://dx.doi.org/10.1145/2858036.2858169

[10] Nielsen, Jacob. (1995). Severity Ratings for Usability Problems. At: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

[11] Oakley, I., and Lee, D. 2014. Interaction on the Edge: Offset Sensing for Small Devices. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (CHI 2014), 169–178. DOI=http://dx.doi.org/10.1145/2556288.2557138

[12] Perrault, S. T., Lecolinet, Eagan, J., and Guiard, Y. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In *Proceedings of the 31st SIGCHI Conference on Human Factors in Computing Systems* (CHI 2013), 1451–1460. DOI=http://doi.org/10.1145/2470654.2466192

[13] Pizza, S., Brown, B., McMillan, D., and Lampinen, A. 2016. Smartwatch in *vivo*. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI 2016). ACM, New York, NY, USA, 5456-5469. DOI=http://dx.doi.org/10.1145/2858036.2858522

[14] Rawassizadeh, R., Price, B. A., and Petre, M. 2014. Wearables: Has the Age of Smartwatches Finally Arrived?. *Communications of the ACM* 58, 1, 45–47. DOI=http://doi.org/10.1145/2629633

[15] Rekimoto, J. 2001. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers* (ISWC'01). IEEE Computer Society, Washington, DC, USA, 21-27.

[16] Schilit, B., Adams, N., and Want, R. 1994. Context-Aware Computing Applications. In *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications* (WMCSA '94). IEEE Computer Society, Washington, DC, USA, 85-90. DOI=http://dx.doi.org/10.1109/WMCSA.1994.16

[17] Smailagic, A., and Siewiorek, D. 2002. Application design for wearable and context-aware computers. IEEE Pervasive Computing. http://doi.org/10.1109/MPRV.2002.1158275

[18] Xu, C., and Lyons, K. 2015. Shimmering Smartwatches: Exploring the Smartwatch Design Space. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '15). ACM, New York, NY, USA, 69-76. DOI=10.1145/2677199.2680599