

Wearable devices for health and wellbeing: Design Insights from Twitter

Omar El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Tareq Nasrallah
Dakota State University
tareq.nasrallah@dsu.edu

Ahmed El Noshokaty
Northern Michigan University
aelnosho@nmu.edu

Abstract

Wearable devices are increasingly recognized for their potential to improve health and wellbeing. However, challenges remain for wide-scale adoption and use. This paper explores perception and reactions towards wearable devices with a particular emphasis on factors that influence the adoption and use to improve health and well-being and which can also inform their design as components of a behavioral change system.

We use social media analytics to analyze and categorize tweets related to major manufacturers of consumer wearable devices from June 1, 2017 – May 31, 2018. We used extant literature on the design of persuasive systems to inform the definition of pertinent categories.

The findings confirmed the relevance of persuasive design features such as Dialog, credibility, and social support, though to various degrees. The analysis sheds light on other user priorities pertaining to device characteristics, integration with other systems, issues surrounding actually wearing these devices on a regular basis.

1. Introduction

Wearable devices such as wristbands and smartwatches are receiving much attention particularly for their potential in improving health and well-being. Individuals are able to connect their phones and other devices to their wearables to track a variety of metrics including steps, calories, heart rate, and sleep (to name a few).

While wearables are currently more likely to be owned by individuals who already seek a healthy lifestyle and want to track their achievements [1], interest continue to increase in expanding the reach of wearables. Moreover, issues remain regarding the sustained use of these wearables. According to Ledger (2014) more than half of U.S. consumers who have owned a modern activity tracker no longer use it [2]. One third of U.S. consumers who have owned an

activity tracker stopped using the device within six months of receiving it. Despite the growing number of wearables' brands, models, and versions available, designing the interaction and interfaces for wearables is still challenging [3]. Accordingly, understanding the users' perceptions toward wearables design and functionality can help to get insights to improve the user interaction with wearable technologies.

In that regard, the everyday usage of social media provides new opportunities for analyzing several aspects of, and patterns in communication. For example, social media data can be analyzed to gain insights into issues, trends, influential actors and other kinds of information [4]. Several studies have taken social media as a rich source for data analysis. Golder and Macy (2011) researchers analyzed Twitter data to study how people's mood changes with time of day, weekday and season [5]. In the area of Information Systems (IS), social media data were analyzed to investigate questions such as the influence of the social network position on information diffusion [6].

The purpose of this study is to leverage social media (Twitter in particular) to examine how the users' experience can help in getting insights to improve the design of wearables devices. The findings provide a foundation for future research into the design of wearables, usability, and issues affecting sustained engagement (a key success criteria for success for wearables in general, and for health and well-being applications in particular). From a practical perspective, the research provides actionable design recommendations and considerations for the design of wearables.

The remainder of the paper is organized as follows: the next section provides a brief literature review followed by a detailed description of the research design and methodology including data collection and social media analytics. The results section summarizes the findings from an analysis point of view that aims to evaluate the general sentiment towards wearable as a backdrop for this research, and as a baseline for future work. The following section provides a detailed presentation of each of the categories in question and implications for design and future research. The paper

concludes with a summary of findings and contributions, and a discussion of limitations and future research.

2. Literature review

The growing of the number of users and popularity of social media as a network for sharing and debating scientific information raises concerns about the types of discussions that surround emerging technologies [7]. As one of the emerging technologies, wearables have become popular in several application domains, including healthcare, entertainment, and others [8].

Wearables are designed and presented to individuals to make them able to quantify and monitor their lifestyles in a systematic manner. This increases awareness among wearable users regarding their health and wellness, and encouraging them to make healthy changes in their lifestyles [9].

Wearable technologies have exclusive benefits within health and wellness (e.g. Using wearables could be a means to assist the young elderly to stay active longer and remain healthy) [10]. Wearable devices and fitness tracking services have become more present, however still suffer from high abandonment rates [11]. Further, some studies indicate that overuse of these devices could trigger depression, anxiety, and obsessive-compulsive disorder and negatively affect mental and emotional health [12]. Accordingly, in the last decade, there is an increasing research interest into the design of wearables as persuasive and behavioral change support systems (BCSS), e.g., [13]. Other research indicates the designs of fitness trackers and wearables designs should concentrate on motivating users and providing support and encouragement regarding the user's progress [11].

With the proliferation of social media, emerging research aims to leverage this media as a rich data source to gain insight into 'real-life' use experiences, aka in the 'wild'. For example, In the context of mobile apps for diabetes self-management, Al-Rahmani et al. (2017) utilize text mining of online user reviews to infer design principles for BCSS [14]. For wearables, Motti and Caine (2016) coded and analyzed 545 wearable users' reviews from Amazon to understand the impact of context in user interaction with these devices [15]. Lowens et al. (2015) relied on 1,349 comments by users of wearable devices analyze users' perception regarding their interaction with wearables [3]. The aforementioned research affirms the potential for analyzing users' posts on social media as a mechanism to better understand their needs and perceptions toward wearables.

While, user reviews of apps and devices can provide insight into users experiences, in the context of wearables, we are equally interested in users' experiences and 'gut' reactions as they deal with these devices on a daily basis. Hence, this research relies on Twitter as a microblogging platform and relies on automated content analysis to maximize exposure to social media data.

3. Research design and methodology

To study users' experience and feedbacks about using wearables, we leveraged Crimson Hexagon, a social media analytics for data collection and analysis [16]. Crimson Hexagon (CH), a social media analytics company, employing an unsupervised and supervised machine learning techniques and text analysis model developed by Daniel Hopkins and Gary King [17]. In general, the steps proceed as follows: First, data collection: This is where the user determines the date range of interest, the social media data sources, the keywords to use to search for relevant posts, and the restrictions to impose (language, geographic location...etc.) Second, analytics: This is where the user selects the appropriate approach (using predefined categories for sentiment or opinion analysis, or user-defined categories). For the latter, the user identifies pertinent categories and a labeled data set for training purposes.

3.1. Data collection

We collected tweets that correspond to representative leading brands of wearables in the US. Those brands are Apple watch, Fitbit, Samsung, Nokia, Garmin, and Pebble. The collected tweets are all selected based on the criteria of having at least one brand name keyword matches set, in our case, the keywords set where the brands' names. We excluded retweets, addresses, and certain words that are not context relevant as shown in Figure 1.

```
(
  (fitbit)
  OR
  (Garmin AND (vivactive OR vivosmart OR vivofit OR forerunner
    OR vivoki OR fenix OR vivomove))
  OR
  (nokia AND (cardio OR bpm OR go OR steel or body))
  OR
  (samsung AND gear AND -"gear VR")
  OR
  (pebble AND (classic OR steel OR time)
    AND -shoe AND -(mountain OR mountains) AND -beach)
  OR
  (("apple watch" OR applewatch OR applewatches)
    AND -(TV OR appltv OR show))
)
AND - (RT OR http OR https)
```

Figure 1. Search query

3.2. Data analysis

Crimson Hexagon employs the ReadMe algorithm developed by Daniel Hopkins and Gary King [17]. This is a supervised learning algorithm that expects the researcher to hand-code a ‘training set’ of documents (posts) into a set of predefined categories. Crimson Hexagon provides an already ‘trained’ model for sentiment and opinion mining, or an opportunity for the researcher to train their own model using user-defined categories.

The ReadMe algorithm is particularly suited when the objective is to know the proportion of the population of posts that fit in specific categories. Rather than calculating this proportions based on the categorization of individual posts, ReadMe gives approximately unbiased estimates of category proportions even when the optimal classifier performs poorly [17].

The key advantage of using a social media analytics platform such as Crimson Hexagon is that it provides access to the “Twitter fire hose”, i.e., it provides access to every public tweet ever posted on Twitter in any language and from any geographic location that meets the search criteria. While it provides the possibility of downloading data for further analysis and exploration, a limitation of Crimson Hexagon is the constraints imposed (mostly by Twitter) on the amount of data the researcher can download. We have addressed this limitation by manually reading and verifying thousands of tweets.

In this research, we use the ReadMe (provided by Crimson Hexagon) to analyze the proportion of tweets that fall into specific categories. We initially utilized Crimson’s ‘built-in’ categories and associated ‘trained’ algorithm to explore the general sentiment and opinion surrounding the use of these wearables. Sentiment analysis uses pre-defined sentiment categories and a large set of training posts (over 500,000) that were hand-labeled as positive, negative or neutral. Crimson Hexagon uses these labeled posts to calculate the frequency distribution of each word, negated word, etc. present in those posts across the positive, negative, and neutral categories. The frequency distributions are then used to construct a model that analyzes each post and classifies its sentiment.

We then trained a separate model to identify the proportion of tweets falling into customized categories reflecting a selected device and a design consideration. The categories are primarily drawn from the literature pertaining to the design of persuasive and behavioral change systems. Most notably following Kukkonen and Harjumaa (2009) [18], we identified dialog

support, system credibility support, and social support. In addition, we included integration support identified in Al-Ramahi et al. (2017) [14] and added a number of categories that are likely to influence adoption and use and/or can have design implications. Examples include cost (affordability), battery and charging considerations, and aspects related to wearing these devices for extended periods of time. Appendix 1 describes each of the categories, keywords delineating each of these categories, and a representative tweet. Using Appendix A as a code book, we manually labeled and distributed 980 tweets over the 13 categories. The training was an iterative process ensuring that each category is clearly outlined by the examples.

4. Analysis results

Over a period of year (Jun-01-2017 to May-31-2018) we collected 868,820 English-language tweets. Figure 2 shows the number of the tweets for each of the representative brands. Overall, 57% of the tweets included gender information while 18% included age information distributed as shown in Figure 3.

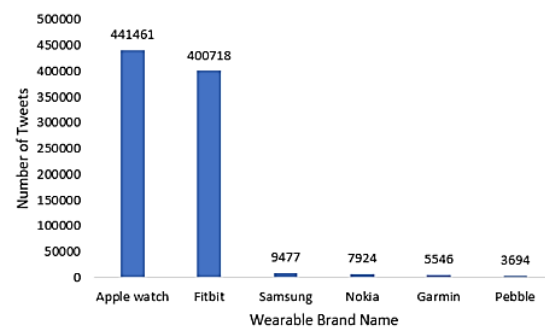


Figure 2. Number of tweets per brand

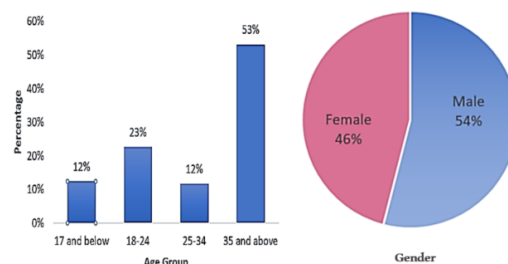


Figure 3. Demographics of user tweets

4.1. Sentiment analysis

Sentiment analysis is an emerging area of Natural Language Processing (NLP) with research extending from document level characterization to taking in the boundary of words and phrases [19]. In addition, the emotion analysis feature provides an additional layer of contextual analysis. Utilizing the "Ekman 6"

(Anger, Fear, Disgust, Joy, Surprise, and Sadness) basic human emotions [20]. We conducted a sentiment and emotion analysis on the collected tweets.

Figure 4 summarizes the results of both analyses for all the collected tweets.

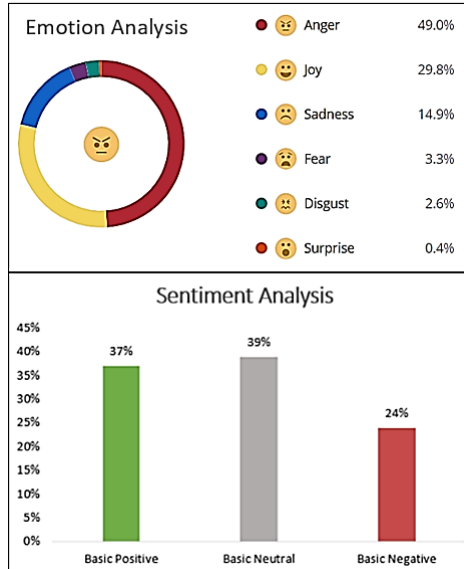


Figure 4. Sentiment and emotion analysis

The emotion analysis results are showing 49% of the tweets classified as anger emotion. This means about half of the tweets datasets are expressing the anger emotion of the users. To better understanding the nature of these tweets in the anger emotion, we explored the anger emotion tweets. Table 1 lists representative tweets. Sentiment Analysis (Positive/Negative/Neutral) and Emotion Analysis are separate classifiers operating independently. There may be cases where Sentiment Analysis has labeled a post as negative and Emotion Analysis has labeled it as joy. For example, the post "I love donuts!! Why do they have to be so bad for me!?" could have a 'Negative' Sentiment Analysis but a 'Joy' Emotion Analysis.

Table 1. Examples of the anger tweets

Why has my @fitbit app started saying "One point oh miles"? Drives me bonkers. Why can't zero be zero? #petpeeve
Trying to re-sync my Apple Watch is becoming a real hassle! 😡
Texting on my Apple Watch is very annoying
apple watch is pissing me off i ain't getting none of my text to my phone
Soooo frustrating doing a full workout when your Fitbit doesn't record any of it 😡 😡
Fitbit friends - anyone else noticing issues with it syncing to your app? Mine keeps struggling with my iPhone for some reason.

4.2. Device and design consideration

Figure 5 is a summary of the proportion of tweets falling into the various categories and the percentage change number of tweets for each category over the analysis period June 1, 2017 – May 31, 2018. For example, 'Credibility support' amounted to 36% of the tweets and its volume have dropped 16% over the analysis period.

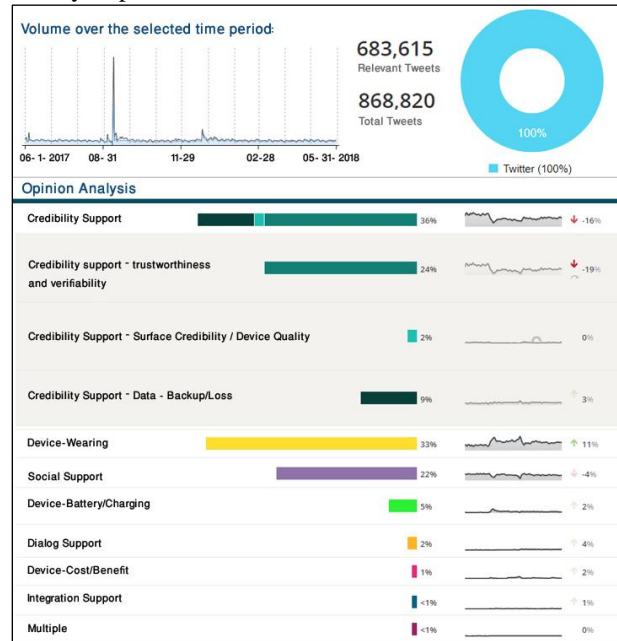
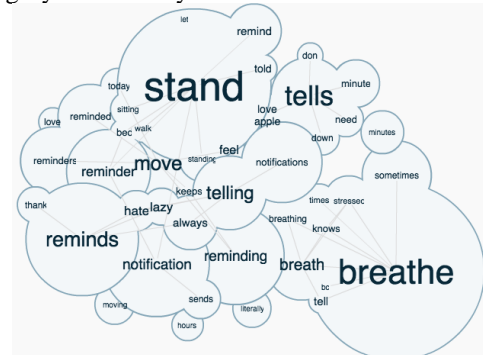


Figure 5. Proportion of tweets by category

Overall, the results demonstrate the identified categories account for 78% of the total number of posts. There are a number of posts that are particularly prevalent, most notably, Device-Wearing, Device credibility (accuracy), Social support, and Data-Backup/loss. In essence, posts pertaining to credibility (perceived accuracy of the devices in measuring activity sleep, heart rate...etc., overall feel of built quality, and concerns/complains of data loss) accounted for 36% of the tweet volume with 24% related to accuracy. Tweets related to the actual wearing of the devices amounted to one third of the tweet volume while social support which is a critical element in persuasive systems accounted for 22%. Battery and charging considerations amounted for 5%, while dialog support (another element in key design consideration for persuasive systems) was responsible for a mere 2%. Figure 6 provides a high-level view of keyword clusters and their relations using a sample of 1,000 tweets. Overall, two clusters relate to the Apple Watch and Fitbit devices reflecting the prominence of the tweets concerning these devices. Battery life, cost,

Dialog support related tweets accounted for only 2% of the tweets. As shown in Figure 7, tweets related to ‘stand’ and ‘breath’ tended to dominate this category followed by ‘move’ reminders.



Manual inspection of the posts are not necessarily positive and may include sarcasm, e.g.,:

“@fitbit dumped you for a Samsung gear frontier s3 for drink water reminder; no more you telling me to post to your suggested features page”

“May I suggest adding a function? “I am in hospital, it’s useless reminding me my 250 steps all day long, I will do my best later””

Manual inspection of the tweets further confirms these results as shown from the following representative examples:

“Apple watch should let us challenge our friends like fit bit does”

Manual inspection of the tweets supports these observations. For example:

“my fitbit registers a step every time my arms goes to hit a key at the piano. RIP step accuracy. hello measurement of practice.”

Credibility support – Surface credibility reflect initial assessment of the credibility of the device and accounted for 2% of the tweet volume. In the context of this research, this category reflects the perception and feel of the overall quality of the device as reflected in Figure 10.

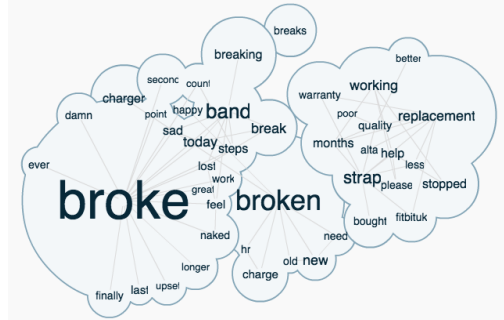


Figure 10. Cluster of keywords for System Credibility – surface credibility and verifiability

Manually inspecting the affirms the importance of quality with a mix of positive and negative experiences. Further, a number of comments pertained to issues with the associated strap/band. Examples of such tweets:

"Thank you, I have a Garmin Forerunner 225 & use Garmin Connect alongside Strava. Never had an issue with reliability thankfully."

“although Apple Watch has much better build quality so it should last on my clumsy paw much longer than fitbits do”

Credibility support – Data loss/back accounted for almost 10% of the tweet volume reflecting the relative importance of this category. This category can fall under Kukkonen and Harjuma (2008) trustworthiness design principle further amplifying the importance of designing wearable devices and supporting ecosystems that garner the trust of the end-user. Figure 11 shows representative keywords and their associations.

As expected, data related tweets are dominant. Concern related to deleted and lost data are also apparent. Manual inspection of the tweets provided further insights into the importance of safeguarding the data collected by these devices, e.g.,

“While doing a swimming workout I wanted to check my heart rate. Apple Watch said ok, displayed the heart rate and deleted the whole swimming session, all the data. So magical.”

“Just had to reset my @Apple Watch which means I've lost allllll of my workout data. So sad...”

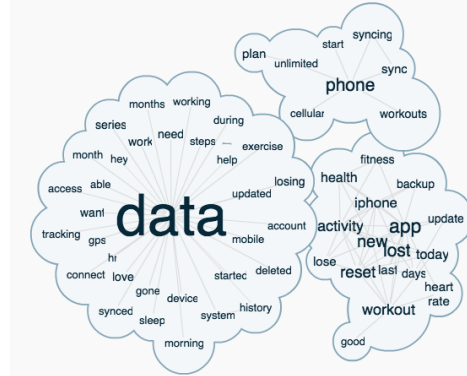


Figure 11. Cluster of keywords for Device – Data loss/backup

Interestingly, device integration category accounted to less than 1% of the tweet volume. Figure 12 Prominent keywords. Examples of these tweets are:

“another some sort of update on your end ... and now my #Charge2 won't sync or register activity; VERY disappointing!!! ☹”

"The gym equipment integration for Apple Watch is siiiiick"

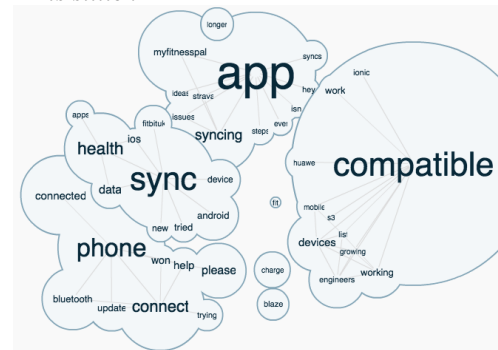


Figure 12. Cluster of keywords for Integration Support

Tweets pertaining to wearing these devices amounted to almost one third of the tweet volume. As the single largest category, this reflects a variety of factors related to wear (Figure 13). These factors range from aesthetics and form factor, to outright issues associated with wearing these devices such as feeling too tight on the wrist or generating a skin reaction or a burning sensation.

Manually inspecting relevant posts in this category further confirmed these observations. Examples of such tweets are:

"I'm torn..I wanna get a Fitbit bc I love data but I'm not used to wearing stuff on my wrist / they kinda look ugly ???"

"Pls do not talk to me about fashion when u wear your Apple Watch with a dress"

Device battery/Charging related tweets accounted for 5% of the relevant tweet volume. Tweets highlighted the importance of battery life to end-users (Figure 14). The tweets peaked slightly surrounding the introduction of the Apple Watch 3 as battery life has been a major concern for the earlier versions of the Apple Watch. Examples of such tweets:

[illegible]

Device cost versus benefit and overall affordability captured only 1% of the relevant tweet volume. As with battery-related tweets, tweets spiked surrounding the introduction of the Apple Watch 3. Figure 15 depicts prominent keywords and their relations. Manually inspecting relevant tweets verified that cost is a concern as shown from these examples:

In this section we discuss and provide a set of design recommendations for wearables upon the analysis results for each category one by one. Overall, while Dialog Support is a critical component of persuasive and behavioral change systems, the mere presence of such support is not necessarily sufficient for successful adoption. Concerns about the frequency, the nature, and the lack of context awareness of notifications and reminders appears to have a counter effect. While context awareness has been recognized in Lowens et al. (2015) [3], it is also apparent that there is a need to personalize the dialog support to the individual user. This leads to the following design and research recommendation:

Future research can investigate how the dialog needs for various users may vary, e.g., by gender, age, health status, physical ability...etc. and can explore ways to optimize the notifications using data collected from these devices.

Future research may study other principles identified in Kukkonen and Harjuma (2008) such as social learning and social facilitation.

Credibility support – trustworthiness and verifiability category's results are consistent with Kukkonen and Harjuma (2008) corresponding design principle pertaining to the importance of trustworthiness and verifiability in persuasive systems. It further confirms a similar assertion by Lowens et al (2015). In essence, user do care about device accuracy and want to trust their devices, however, despite recent advances in the technology, device accuracy, and thus device credibility remains a concern. Accordingly,

DR3: Device accuracy is a user expectation and design requirements. Devices need to ensure more accurate measurement and calibration algorithms.

Credibility support – Surface credibility, despite the relatively low volume of tweets during the analysis period, quality is still a concern that extends beyond the device. Hence:

DR4: Perceived quality of devices is a non-functional design requirement. As devices intended for wearing extended periods of time, wearables should be designed with build quality as a priority and should include the device, as well as its accessories and supporting ecosystem.

Credibility support – Data loss/back, while backup policies and practices are standard for most businesses, it is not common practice for end-users of information technology. Hence,

DR5: Wearable devices and supporting ecosystems need to include failsafe mechanisms for securing data collected by these devices even as the device itself fails and require reset.

Device integration category, the relatively small volume in this category is somewhat surprising given its prominence in earlier research, Al-Ramahi et al (2017). One possible explanation is that this category can serve as a supporting category related to transferring and syncing data for sharing (social support), backup, and a variety of other functions. Regardless, manual inspection of the tweets. While further research is needed, one can postulate the following:

DR6: Wearable devices need to include provisions for the seamless, reliable, and open integration with other devices and related ecosystems, e.g., Apple and Samsung Health.

By extension, one can postulate that:

DR7: Existing and future health and well-being ecosystems and platforms need to adopt an

open application program interface (API) to allow seamless integration of various wearable devices in support of key persuasive and BCSS functions such as Social Support.

In reference to the device – wearing category, it is evident that as wearables, non-functional related to wearing these devices for extended periods of time and in a sense interfering with 'every-day-life' consideration such as look and appeal poses additional design requirements. It is simply not enough for these devices to have the necessary functionality and exhibit high quality. Wearables devices are more than technical gadgets. Hence,

DR8: Wearable devices design should account for non-functional requirements related to users' expectations for look, appeal, and form factor.

The aforementioned design requirements as wearables continue to evolve in various forms (rings, bracelets, clothes accessories, etc.). The key is how these devices can blend in with the users' look, or at least not interfere with it.

Evidently, the idea and expectations to charge yet another device (besides the mobile phone) appeared a concern. Another issue reflected in the tweets is interfering with other functions, e.g., the need to charge, and the need to wear to track sleep. Forgetting to charge, forgetting the charger or not having access to a charge is another concern that echoed repeatedly in the tweets. It is worth-while noting that a number of tweets even suggested the use of kinetic energy to eliminate the need for manually charging these devices. Hence:

DR9: Wearables should include mechanisms to reduce (if not eliminate) the burden for having to regularly charge these devices.

Drilling into the cost-related tweets reveals that for the tweets that have identifiable age, almost 60% are 24 years or younger. This may explain the lower volume that may possibly allude to cost/price being less of an issue for the older age groups (>25).

6. Conclusion

In this research we examine user experiences and reactions to wearable devices to inform design, future research and development. Using supervised learning and sentiment analysis techniques we followed an automated content analysis methodology to analyze 868,820 tweets spanning an entire year for major wearable brands. The analysis identified nine design requirements and related research priorities (Table 2). Further, the methodology could be applied to other IoT products by extracting pertinent user insights that can guide research and development to improve their design and functionality.

Table 2. Summary of design recommendations

DR1: Dialog support should be personalized to the individual user and contextualized to situation.
DR2: BCSS systems should seamlessly incorporate various social support principle.
DR3: Devices need to ensure more accurate measurement and calibration algorithms.
DR4: Wearables should be designed with build quality as a priority.
DR5: Wearable devices and supporting ecosystems need to include failsafe mechanisms for securing.
DR6: Wearable devices need to support seamless, reliable, and open integration with other devices.
DR7: Platforms need to adopt an open application program interface (API) to allow seamless integration of various wearable devices.
DR8: Wearable devices design should account for non-functional requirements related to users' expectations for look, appeal, and form factor.
DR9: Wearables should reduce (if not eliminate) the burden for having to regularly charge these devices.

While the design recommendations can inform future development of these devices and their supporting ecosystems, the findings highlight that warrants future research. The research limitations and possibilities for improvement include: Firstly, the further refinement of the categories, possibly focusing on specific design considerations, e.g., user interaction. Secondly, augmenting this research with surveys of users to better understand specific concerns or areas for improvement.

7. References

- [1] L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson, "The Rise of Consumer Health Wearables: Promises and Barriers," *PLOS Med.*, vol. 13, no. 2, p. e1001953, Feb. 2016.
- [2] D. Ledger and D. McCaffrey, "Inside Wearables-Part 1: How the Science of Human Behaviour Change Offers the Secret to Long-Term Engagement," 2014.
- [3] B. Lowens, V. Motti, and K. Caine, "Design recommendations to improve the user interaction with wrist worn devices," 2015, pp. 562–567.
- [4] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics – Challenges in topic discovery, data collection, and data preparation," *Int. J. Inf. Manag.*, vol. 39, pp. 156–168, Apr. 2018.
- [5] S. A. Golder and M. W. Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures," *Science*, vol. 333, no. 6051, pp. 1878–1881, Sep. 2011.
- [6] A. Susarla, J.-H. Oh, and Y. Tan, "Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube," *Inf. Syst. Res.*, vol. 23, no. 1, pp. 23–41, Mar. 2012.
- [7] K. K. Runge et al., "Tweeting nano: how public discourses about nanotechnology develop in social media environments," *J. Nanoparticle Res.*, vol. 15, no. 1, Jan. 2013.
- [8] V. G. Motti and K. Caine, "Users' Privacy Concerns About Wearables," in *Financial Cryptography and Data Security*, vol. 8976, M. Brenner, N. Christin, B. Johnson, and K. Rohloff, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 231–244.
- [9] M. U. Warraich, "WELLNESS ROUTINES WITH WEARABLE ACTIVITY TRACKERS: A SYSTEMATIC REVIEW," p. 14.
- [10] C. Walden and A. Sell, "Wearables and Wellness for the Young Elderly - Transforming Everyday Lives?," 2017, pp. 637–650.
- [11] K. Chen, M. Zdorova, and D. Nathan-Roberts, "Implications of Wearables, Fitness Tracking Services, and Quantified Self on Healthcare," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 61, no. 1, pp. 1066–1070, Sep. 2017.
- [12] Q. Wu, K. Sum, and D. Nathan-Roberts, "How Fitness Trackers Facilitate Health Behavior Change," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 60, no. 1, pp. 1068–1072, Sep. 2016.
- [13] K. Mercer, M. Li, L. Giangregorio, C. Burns, and K. Grindrod, "Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis," *JMIR MHealth UHealth*, vol. 4, no. 2, p. e40, Apr. 2016.
- [14] M. A. Al-Ramahi, J. Liu, and O. F. El-Gayar, "Discovering Design Principles for Health Behavioral Change Support Systems: A Text Mining Approach," *ACM Trans. Manag. Inf. Syst.*, vol. 8, no. 2–3, pp. 1–24, Jun. 2017.
- [15] V. G. Motti and K. Caine, "Smart Wearables or Dumb Wearables?: Understanding how Context Impacts the UX in Wrist Worn Interaction," in *SIGDOC '16*, September 23 - 24, 2016, Silver Spring, MD, USA, 2016, pp. 1–10.
- [16] C. Hexagon, "AI-Powered Consumer Insights Company | Crimson Hexagon." [Online]. Available: <https://www.crimsonhexagon.com/>. [Accessed: 14-Jun-2018].
- [17] D. J. Hopkins and G. King, "A Method of Automated Nonparametric Content Analysis for Social Science," *Am. J. Polit. Sci.*, vol. 54, no. 1, pp. 229–247, Jan. 2010.
- [18] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive Systems Design: Key Issues, Process Model, and System Features," *Commun. Assoc. Information Syst.*, vol. 24, p. 18, 2009.
- [19] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R," *Inform. Econ.*, vol. 22, no. 1/2018, pp. 25–38, Mar. 2018.
- [20] P. Ekman, "Facial expression and emotion," *Am. Psychol.*, vol. 48, no. 4, pp. 384–392, 1993.

Appendix A: Codebook for labeling categories

	Category	Description	Keywords	Examples
1	Dialog support	Captures system interactivity aiming at providing some degree of system feedback to its users, potentially via verbal information or other kinds of summaries. This includes dialog that helps users keep moving towards their goal or target behavior (Kukkonen and Harjumah, 2008).	reminders, notifications, praise, rewards, ...etc.	I wish I could customize the "reminders to move" on my #fitbit. I respond better to a reward system. "get up now and you can have a cookie"
2	Social support	Emphasizes how the device motivates users by leveraging social influence such as social facilitation, social comparison, competition, and recognition (Kukkonen and Harjumah, 2008).	social, sharing, Twitter, Facebook, Instagram	I have all of 3 friends for my Apple Watch sharing. I wish you could connect to Facebook or something like Fitbit.
3	Credibility support – trustworthiness and verifiability	Focuses primarily on device sensor accuracy and the trustworthiness and verifiability of the measurements provided by the device (Kukkonen and Harjumah, 2008).	Accuracy, measurement, estimate, overestimate, inaccurate	Anyone else's Apple Watch terribly inaccurate with exercise data?
4	Credibility support – surface credibility/ device quality	Accounts for the overall competency of the device's look and feel, in addition to reliability and functional quality (Kukkonen and Harjumah, 2008).	Quality, reliability, "does not work", "not working", broke	Really disappointed with the quality of my 2nd #fitbit in a year! What other options are there? @fitbit get it together
5	Credibility support – Data-loss/backup	Captures issues with losing users' data and workout and measurement history.	Reset, restore, backup, back up, lost, workout data, history	All that workout activity is gone just like that... I'm so pissed that I feel like breaking my @Apple watch ☹
6	Integration support	Reflects those topics pertaining to connectivity, compatibility, and integration with mobile devices, cloud services, mobile phone apps, and medical devices (Al-Rahmani et al., 2017).	Connectivity, connect, link, Bluetooth, syncing, sync, Compatibility, Integration	The hardest test I've encountered so far is trying to figure out why my fitbit won't sync
7	Device-wearing	Relates to various issues, and concerns surrounding the wearing and everyday use of the wearable devices	Wearing, comfort, aesthetics, form factor, fashion, look, appeal, tight, burn	@GarminFitness @Garmin I've got a burn on my wrist from the optical sensor on my ForeRunner 35. I've tried messaging but with no response.
8	Device – Battery/charge	Concerns about battery life and charging issues.	Battery, charge, energy, died," battery life"	My Fitbit isn't even a year old and the battery power sucks now!! :(
9	Device-cost/benefit	Pertains to the price and the cost of the device, particularly taking into account the perceived benefits.	Cost, pay, price, money, benefits, value for money	Disappointed that @fitbit chose to discontinue the ONE. But not enough to pay 3x's the price on ebay. Looking for recc's on another similar #fitnesstracker. Preferably NOT a wrist-worn.
10	Multiple categories	The post that contains keywords and concerns that relates to two or more of the above Categories		i love my fitbit. it does steps taken, calories burned, you can log your workouts, track how much water you drink, syncs with myfitnesspal app so i can see calories consumed also. also does heart rate, miles walked, active minutes, and tracks my sleep.
11	Irrelevant (off-topic)	The post that has no related content for any of the above categories		Can Christmas come quicker bc damn all I want is my phone & Apple Watch ☹