

Crowdsourcing Queue Estimations in Situ

Jorge Goncalves¹, Hannu Kukka², Iván Sánchez³, Vassilis Kostakos¹

¹Community Imaging Group, University of Oulu, Finland

²Urban Computing and Cultures Group, University of Oulu, Finland

³Interactive Spaces Group, University of Oulu, Finland

firstname.lastname@ee.oulu.fi

ABSTRACT

We present the development and evaluation of a situated crowdsourcing mechanism that estimates queue length in real time. The system relies on public interactive kiosks to collect human estimations about their queue waiting time. The system has been designed as a standalone tool that can be retrospectively embedded in a variety of locations without interfacing with billing or customer systems. An initial study was conducted in order to determine whether people who just joined the queue would differ in their estimates from people who were at the front of the queue. We then present our system's evaluation in four different restaurants over 19 weekdays. Our analysis shows how our system is perceived by users, and we develop 2 ways to optimise the waiting time estimation: by correcting the estimations based on the position of the input mechanism, and by changing the sliding window considered inputs to provide better prediction. Our analysis shows that approximately 7% of restaurant customers provided estimations, but even so our system can provide predictions with up to 2 minute mean absolute error.

Author Keywords

Crowdsourcing; queuing; waiting time; estimations; situated; restaurants; tablets; public displays.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Queues are logistic mechanisms in which a group of individuals wait in order, using a first-in first-out procedure, generally with the goal of obtaining a good or service [23]. Studies have repeatedly shown that waiting time has a substantial impact on the perception of that service [33], and therefore individual establishments are motivated to reduce their waiting times and/or providing an enhanced

queuing experience. This is especially true in environments where a single service provider owns multiple service points that are not in direct competition with each other over customer volumes. Examples of such are hospitals, universities, industrial complexes, etc. where one operator is often responsible for providing restaurant services for the entire establishment. In such settings, both customers and service providers will benefit from an approach such as the one presented here, since it can help distribute the customer load more evenly between locations.

Existing queue prediction tools such as those utilising cameras [1], sensors deployed in the environment [2] or WiFi and Bluetooth signals coming from queuing individuals' mobile devices [4,36], while potentially accurate have a number of drawbacks. For instance, while systems that utilise WiFi signals have reported the lowest estimation error [36], they have three important drawbacks. First, they require that a large number of people in the queue have a specific application installed in their mobile devices in order to contribute data to the crowd-sensing system. Second, they usually assume that people automatically join a queue when they enter a particular premise. They do not consider situations in which people just meet other individuals that are, for example, already having coffee or lunch without participating in the queue. Third, they do not consider multi-queuing environments.

We argue that a user-centric approach involving situated crowdsourcing kiosks can assist in mitigating these drawbacks: customers can generate and share queue information amongst themselves with low effort and, as a by-product, also help the service points to provide a better experience for the customers. The main objective of our research is to test the above assertion: can crowdsourcing provide a method for generating reliable waiting time estimates for services in close physical proximity? Would individuals be willing to contribute to such a crowdsourcing service, and how would they perceive its usefulness? Would they provide reliable data, or would they vent their frustration by providing erroneous data?

To answer these questions, we designed a study using four in-campus restaurants as research case studies. The restaurants were retrofitted with public interactive kiosks where customers waiting in the queue could input their estimation of the current waiting time. To determine whether a person's position in the queue affects their

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3592-8/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2818048.2819997>

estimation, we conducted an initial study, which informed us how to calibrate the collected crowdsourcing data to account for individuals' bias. Subsequently, we evaluated the accuracy of the system and its reception by customers during a 19-day deployment.

RELATED WORK

Queuing: Social Aspects, Modelling and Quantifying Techniques

Due to the human factor in queuing, many disciplines have studied such mechanisms. Sociologists and psychologists are mainly concerned with the behaviour and attitude patterns arising in physical queues [24], as well as how different social values affect people's perception in queuing environments. For instance, Larson [21] analyses different psychological and physiological factors that might affect customers' perceptions and attitudes experienced in a queue setting, such as social justice and use of time. Sociologists are also interested in studying this phenomenon to better understand other broader subjects such as crowd behaviour and cultural values regarding public behaviour [32].

On the other hand, business psychologists and marketing researchers have also developed a rich and detailed literature on queues. It has been often reported that the waiting time in a queue affects the customer's evaluation of the quality of a service [5,33]. However, social and psychological factors may influence one's perception of waiting time, and therefore perceived quality of service [28]. Maister [22] enumerates multiple psychological factors such as anxiety, fairness, and social interaction in queues that also effect one's perception of waiting time. In addition, Taylor [32] discusses the effects of filled waiting time on evaluations of service. A substantial number of studies in literature aim to analyse queuing and its implications for quality of service perception in very specific contexts such as supermarkets [3], airports [19], tourist services [9] and tolls [34]. The finding that reducing waiting time improves quality of service is common. Hence, it has become a priority to reliably model the queuing environment so that service providers can estimate, for example, the number of service points necessary to achieve a determined quality of service, thereby balancing cost versus quality of service.

To this end, queuing theory provides the mathematical background in order to create such models, and has its origin in the field of telecommunications. Queuing theory is an extensive field and a useful overview is provided by Gross [15]. It is important to highlight that, as noted by Bulut et al. [4], queuing theory cannot be applied without knowledge of the arrival process, the service distribution time, and the number of servers. For such cases they point to other alternatives for estimating the waiting time, particularly using data mining. Alternatively, in certain cases it is not necessary to predict the behaviour of the queue in the future, and it suffices to simply determine the behaviour queue at the moment. For this purpose it is

possible to adapt methodologies used to measure crowd collectiveness, whereby we effectively consider queues as a particular case of crowd collectiveness [39].

Previous work has utilised several technologies such as wireless sensor technologies [29] or computer vision algorithms [18] to quantify crowd size and queue size. Choi et al. [7] present a method to detect collective activity recognition (talking, queuing) using computer vision algorithms. Zhou et al. [39] use a similar approach, but with emphasis on analysing how individuals move in a crowd. In contrast, Aubert used computer vision algorithms to measure the length of a queue (not the number of individuals) with high accuracy [1]. Finally, Hsieh et al. used a Kinect system above a business' doorway to capture the situation of the pedestrian flow [18]. They report measuring bidirectional flow of people with almost 100% of accuracy in real time.

However, there are certain drawbacks to using computing vision methods to estimate amount of individuals in a queue. First, the cameras need to be placed in optimal positions to minimize algorithm errors. Sometimes, especially for long queues, computer vision systems face occlusion problems requiring multiple cameras for measuring the real length of a queue. In addition, queue cameras raise privacy concerns similar to surveillance cameras. Finally, computer vision algorithms can be complex and computationally intensive, and subject to lighting conditions during the day.

Wireless sensor technologies have also been used to estimate crowd sizes. For instance, RFID has been proposed to measure the length of a supermarket's line [3]. Alternatively, O'Neill et al. [28] tried to predict the crowd density that crossed certain street section and classify the pedestrian flow by a combination of human observation and Bluetooth scanning. In more recent work, Kostakos et al. [20] measured the time a single passenger spent on a bus, and hence collect data about passengers' end-to-end trips. The system detected when a passenger's mobile device was discoverable using a Bluetooth scanner. When the phone is not discoverable anymore, the system can infer that the owner has left the bus. In combination with GPS technology it was possible to know the route a single passenger had taken with high accuracy.

Estimating the queue length in restaurants using Bluetooth technology is quite challenging. On one hand, we must assume that a majority of customers have their Bluetooth transceiver activated. On the other hand, we only can calculate the number of customers that are in the restaurant; we cannot guarantee that the customer is waiting in the queue or just having some coffee.

Crowdsourcing Waiting Times

Crowdsourcing has seldom been reported in the literature in conjunction with estimating queue waiting times. In principle, crowdsourcing relies on gathering contributions

from a large population or a large group of users or consumers. Typically, the objective of crowdsourcing is to divide large amount of work into small tasks that can be performed by individuals [11,16]. Crowdsourcing has been used to create new content collectively [29], for humanitarian aid [10], collecting public opinion [17] and for solving complex problems that either are difficult for machines to solve or would otherwise take too long [12,31].

Some prior work has attempted to measure queue sizes by combining the potential of crowdsourcing with the power of social media [25]. For instance, publications in a social media can be used to geolocate their authors and forward them unsolicited questions regarding to aspects of that particular location (e.g. the weather conditions, or queue length in a nearby cinema). In many ways our work resembles this approach: we use public interactive kiosks to solicit input, rather than users' personal devices. This has the benefit of avoiding unsolicited requests, and at the same time overcomes many privacy concerns, which arise from the fact that one's location may be revealed once they provide an assessment of the queue at a particular establishment. Any customer is able to use (but also to game) this system, since it is not linked to any id or application.

Crowdsourcing to estimate waiting times has been used in conjunction with wireless technologies. For example, Weppner et al. [37] have shown how to estimate the crowd density using Bluetooth and leveraging collaboration between close-by devices. They claim to improve the recognition rate by 30% when compared to just using the absolute number of discovered devices, Furthermore, Bulut et al. [4] developed a system that approximates the queue waiting time in a university cafeteria using indoor positioning methods (WiFi Access Point, GPS and cell tower triangulation). The system measures the time that a person stays in the cafeteria, and the authors use this to approximate the waiting time. Their system is able to predict the actual waiting time using the historical data and using heuristics based on time-series estimation. They managed to reduce mean absolute error to be less than 2-3 minutes. Their underlying assumption is that waiting time is somehow correlated with the time they spent in the cafeteria. To account for this, some follow-up work [5] has made use of devices' accelerometers to determine whether the user is in a queue or not. A similar approach by Wang et al. [36] used WiFi signal strengths from a WiFi monitor located at the service area. Analysing the signal strength patterns of the WiFi signal from restaurant client's mobile phones, authors claim that they are able to measure the queuing waiting time with a maximum estimation error of 10 seconds.

SYSTEM DESCRIPTION

Based on the reviewed literature, there are three main alternatives to improve user's satisfaction in queue environments: careful control of the waiting environment,

promoting social interaction, and reducing queuing time. The first strategy is not feasible for ad-hoc approaches such as our own, and the second beyond our scope. Therefore, we aim to ultimately reduce waiting time by giving reliable information to customers who are trying to decide which restaurant to visit on our campus.

Our system consists of interactive kiosks (Figure 1) that were installed in four restaurants on a university campus. Each kiosk consists of a touch-enabled Android tablet with a 10.1" touch-screen at approximately 1.1 meters from the ground and WiFi connectivity. The software running on the kiosks serves a single purpose: it invites customers to enter their assessment of how long the actual waiting time in the restaurant queue is. Users' inputs are given with a single touch, and are forwarded to a back-end system (Figure 2) which produces a prediction of the queue waiting time in each restaurant. This information is made publicly available through several large displays across the campus, and via a website that users can visit on their mobile phone.



Figure 1. A kiosk that collect crowd estimations (left), and a public display displaying real-time predictions (right).

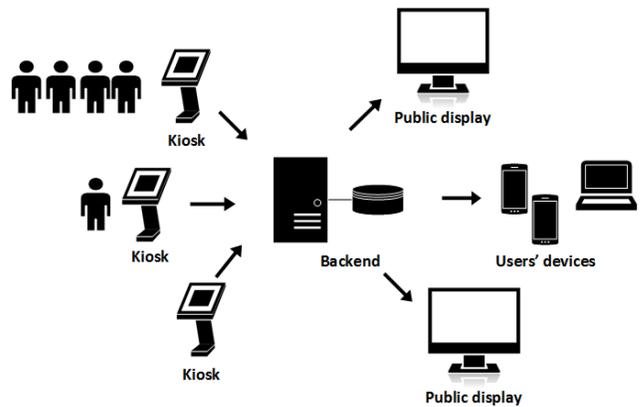


Figure 2. The main components of our system.

Crowd Input and Queue Visualisation

The kiosk application has a single screen with the sole purpose of collecting queue waiting time estimations from restaurant customers (Figure 3). The interface was designed to minimise any affordance of exploration, the tablet's operating system was locked to "kiosk mode", and the physical buttons of the tablet were physically obstructed by the kiosk enclosure. All these decisions were intended to discourage users from appropriating the kiosk otherwise, therefore ensuring that the kiosk did not cause delays to the queue in itself.

Interaction with the tablets is touch-based, and customers waiting in the queue indicate their estimation of the current waiting time using a visual scale (Figure 3 top). An A4 sheet of paper with further instructions was placed below each tablet (Figure 1). These instructions can be seen in Figure 3 bottom. Each touch event is timestamped and converted to a granular number between 0 and 15 based on the screen coordinates of the touch (*i.e.* the scale is granular despite visual appearance). Once user input is collected, the application gives immediate acknowledgement to the user and disables input for 3 seconds to avoid accidental input or abuse. The scale was purposely constrained to minimise explicit input since actually typing a number would take more time, and be prone to errors and misappropriation in a public setting as suggested in other situated crowdsourcing deployments [13,14].

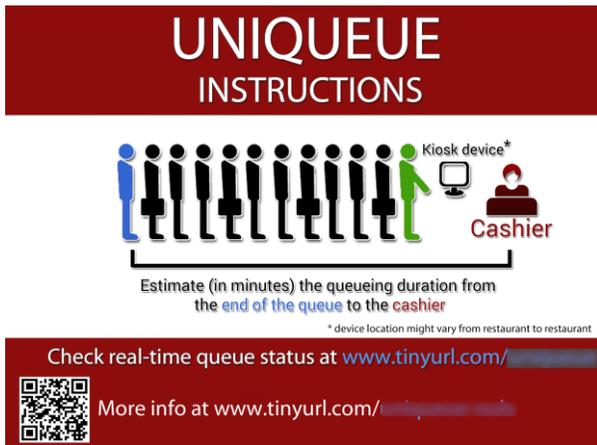


Figure 3. Top: The kiosk app UI. Bottom: A4 sheet with instructions that was place below each tablet.

The web page shown on public displays (and accessible to users' own devices) was built to visualise the queue prediction for all restaurants. It included real time estimates of queueing times, the opening hours, the daily menu of each restaurant, and a graphical and textual representation of the current queue situation (Figure 4). We note that the number of coloured silhouettes either always precisely represents the current queue prediction, or a slight overestimation of it (*e.g.*, if the current estimated queue waiting time is 10 minutes, then 4 silhouettes would appear coloured). This decision was based on literature, which states then when providing queue predictions, one should always meet them or exceed them [27].

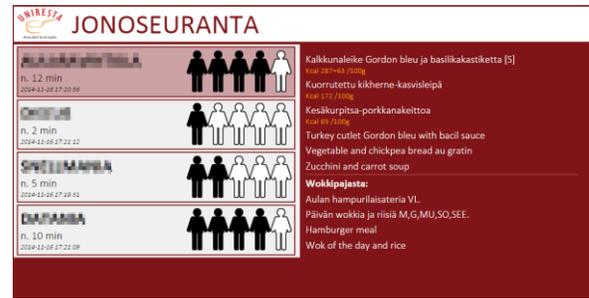


Figure 4. Webpage UI.

Queue Estimation

As Bulut notes [4], we cannot use traditional queue models *a priori* since we do not have an estimation on the arrival rate or service time in the restaurants [15]. Instead, we have a crowdsourced set of queue estimations contributed by users with an irregular sampling rate. The most important parameters that can influence prediction performance are:

- The true arrival rate of customers. This depends on the time of the day, the day of the week, the season and the University's teaching schedule, among others, and is beyond our control.
- The error in customers' estimations (in minutes). This can depend on the customers' attention, the position of the kiosk, and its usability. We assess this factor in Study 1.
- The function used to generate an estimate from the crowdsourcing contributions. We assess this factor in Study 2.

Thus, to actually predict the waiting times from crowdsourcing input, we first need to account for errors and bias in customers' estimations, and then determine an optimum way to convert the crowd's estimates into a prediction. The former is determined in Study 1. For the latter, we opted to apply a weighted average function so that the weight of user estimations decreases over time, placing more emphasis on more recent data. In Study 2, we determine the ideal window size. Equation 1 shows the formulation of the weight function.

$$\frac{\sum_{i=0}^n w_i \cdot C(x_i)}{\sum_{i=0}^n w_i} \quad (1)$$

where $w_i = f(\text{number of close samples, data age})$, and $C(x_i)$ is a correction function.

The Restaurants

The on-campus restaurants where we deployed our system are all characterized as mostly lunch restaurants, serving mainly students and faculty members of the university. All restaurants follow a self-service model, where customers first obtain a tray, flatware, *etc.*, and then proceed to take food from large containers. Customers then pay for their meal and proceed to seat themselves within the restaurant main dining area. Restaurant R1 is the largest of the four (550 seats), located at the heart of the university, and visited

by students from all faculties. Restaurant R2 has 200 seats, and is located near the Faculty of Education. Restaurant R3 has 150 seats and is located near the Faculty of Humanities and a zoological museum that is also open to the general public. Finally, R4 has 125 seats and is located near the Faculty of Technology. All restaurants are closed on weekends.

EVALUATION

We conducted two studies. In Study 1 we evaluate how the placement of the input device affects users' bias in queue estimation. We then conducted in-situ observations and interviews to understand how queues form in the various restaurants on campus. Then, in Study 2 we ran a field trial lasting 19 weekdays, collecting detailed log data of system usage and ground truth data. At the end of the deployment, we interviewed a number of customers to assess their opinions of our system. Finally, a survey was launched to gather additional data regarding the routines of customers (in addition to the interviews and observations in Study 1), as well as provide more insights regarding the use of the system and any changes in behaviour (in addition to the interviews conducted in Study 2).

Study 1: Kiosk Positioning and Estimation Errors

In Study 1 we sought to determine whether kiosk placement (near the front or back of the queue) had an effect on the queue waiting time prediction. We decided against having two kiosks on each restaurant (one for when customer arrives and one for when they pay) to calculate precisely the amount of time spent in the queue as this would require tracking individuals and would likely increase the barriers to contribution. Previous research has highlighted the importance of kiosk placement, for example in healthcare and supermarkets [3,34]. Further, Tom & Lucey [34] demonstrate that the location of a kiosk influences the types of tasks performed on it, the services activated, and the accuracy of the estimations. Therefore, in Study 1 we wanted to establish whether placing the kiosks at the back or front of the queue had an impact on the error in estimation customers made.

Method

A version of the kiosk application was installed on a Samsung Galaxy Tab Pro 8.4" tablet. While carrying this tablet, we asked 42 distinct participants to estimate the current queue waiting time in minutes (*i.e.* what the current situation was, not how long it took them to get serviced) using our interface. Half of the participants were approached at the *back of the queue* (the last person who just joined the queue, $N=21$) while the other half were approached at the *front of the queue* (just after paying the cashier, $N=21$). We avoided having participants make several estimations, as we wanted their selections to be as organic as possible. For each data point we also collected ground truth data manually. In both cases we measured how much time it took from the moment they used our tablet until the last person in the queue (at the time) was eventually served.

The measurements were carried out in the different restaurants during two different periods of the day: during lunch (10:30 till 12:00) and early dinner (15:00 - 16:00). This study was designed to ensure extensive customer population sampling (*i.e.* people from different faculties), and varying queue sizes. One researcher carried the tablet and asked random people in the queue to estimate the waiting time using the application, while simultaneously another researcher measured ground truth values for the waiting time. We made sure that all data was independent by ensuring that one person's answer does not affect another's (e.g. a friend). Waiting time was measured from the time a person arrived to the queue, until s/he had paid at the cashier.

Results

Figure 5 shows the error in estimation (minutes) for participants in the back and front of the queue. An independent-samples t-test indicated that there was a statistically significant difference of estimated error between the two groups when compared to the ground truth ($t(40) = 3.45$, $p < .01$). Participants who provided estimations at the back of queue overestimated the waiting time (error: $M = .89$, $SD = 1.76$) while those at the front of the queue underestimated the waiting time (error: $M = -1.51$, $SD = 2.66$).

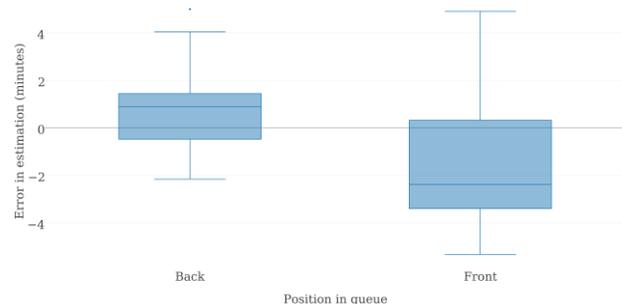


Figure 5. Estimated error (minutes) for participants in the back and front of the queue.

Interviews, Survey and Observations

We conducted in-situ interviews and observations to understand how queues form in all four restaurants on campus (designated R1-R4). We later added data from a survey. We observed the queue formation and dynamics throughout the opening hours of the restaurants (from 10 am till 5 pm), focusing on how people arrived to the restaurants (in groups or alone), and whether certain behaviours were frequent during queuing. We also interviewed 8 students and collected survey answers from an additional 24 participants (21 students, 3 staff) about their lunch-related routines, including restaurant selection, the time they usually have lunch and the social context in which they usually eat. The interviewees were approached at the restaurant, and then moved towards a more secluded area to minimise any disruption, while the survey was distributed via internal mailing lists. No rewards were given to the participants.

Observation data showed that during the morning hours people are more likely to have lunch in small groups (Figure 6). Conversely, in the afternoons we observed more individuals queuing and eating alone. Typically, the observed groups were rather small, up to 4 individuals, although on some occasions we noticed larger groups (up to 10). This was confirmed by our survey results in which 83% of participants reported eating with friends or colleagues. However, it seemed that these groups were created *ad-hoc* in the restaurant when friends or acquaintances happened to meet, rather than purposefully coming together for lunch.

Further, in the interviews and survey, participants reported a wide range of reasons why they chose a particular restaurant. These reasons were socially-driven (their friends are usually there), location-driven (the distance from their classroom to the restaurant), based on the length of the queues or simply based on their everyday routine. The majority of participants reported having lunch between the hours of 11am and 1pm (only one participant reported normally going for lunch after 1pm). While some respondents noted that rush hours are unavoidable, others remarked that they prefer to delay their lunch to avoid queues altogether which ultimately leads to suboptimal eating patterns. In general, participants were quite open and interested in our proposed system, particularly if information about queue predictions in distant restaurants is made available online.



Figure 6. Queuing and ambient at one of the restaurants during lunch time.

Study 2: Field Trial

We deployed the system for 19 weekdays in the four restaurants. The kiosks were placed on the natural path of the queue in each restaurant, typically at the entrance of the restaurant (normally the back of the queue), but the selection of potential locations was naturally constrained by architectural qualities, accessibility issues, or lack of utilities such as power sockets.

Additionally, we placed four 46” public displays (Figure 1, right) throughout the university campus to provide passers-by with up-to-date information on the queuing times in the restaurants. Because we needed to generate queue predictions for the public displays, but did not yet have enough data to determine the ideal window size for Equation 1, we used an arbitrary window of 15 minutes. Thus, we consider the crowdsourced data from each

restaurant for the last 15 minutes and derive a weighted average. Our subsequent analysis shows that this arbitrary choice was not far from an optimal value. We also collected detailed log data from the system, including: input from the kiosks, webpage visits, number of touches on the public displays to access menu information. We also obtained revenue data from each restaurant, collected through their till system. In addition, on 2 days we manually collected ground truth data on queue waiting times for the whole day (10am to 4.30pm).

Quantitative Results

The four kiosks collected a total of 3633 user inputs during the 19-day deployment. The largest restaurant (R1) accounted for 53% of the inputs, R2 for 29%, while R3 and R4 accounted for 9% each (Figure 7). A Pearson’s two-tailed test showed a significant correlation between the estimations by users and volume of inputs for R1 ($r(562) = .19, p < .01$), R2 ($r(298) = .5, p < .01$), R3 ($r(168) = .24, p < .01$), but not R4. This suggests that higher waiting time predictions were associated with more crowdsourcing contributions, which is consistent with our expectations: as queues get longer we expect more input to our system and longer waiting times.

Using the revenue data provided by each restaurant during the field trial, and the average number of estimations in each kiosk collected by the system, we can calculate the kiosk usage rate. This reflects the percentage of the restaurant customers that clicked in the kiosks. Overall, our system was used by 7% of the campus restaurants customers. Table 1 presents usage rate per restaurant, excluding days where there were technical difficulties (e.g., lack of WiFi connectivity). Figure 7 shows the cumulative number of data points collected by each kiosk during deployment.

Figure 8 summarises the intermittency of the data collected by the kiosk: it shows a histogram of the time between subsequent estimations on kiosks for each restaurant. Unsurprisingly, the larger restaurants had lower intermittency between customer inputs. The graphs follow an exponential distribution, which is consistent with the theoretical assumption of a Poisson distribution of arrivals per unit of time [15]. This suggests that the rate of data collection on kiosks is similar to the rate of customers’ arrivals, suggesting that the sampling rate of 7% (Table 1) is consistent over time.

	R1	R2	R3	R4	Total
Average daily customers	1176	655	787	525	3143
Average daily estimations collected	101	57	21	26	205
Usage rate (%)	9%	9%	3%	5%	7%

Table 1. Kiosk use in different restaurants.

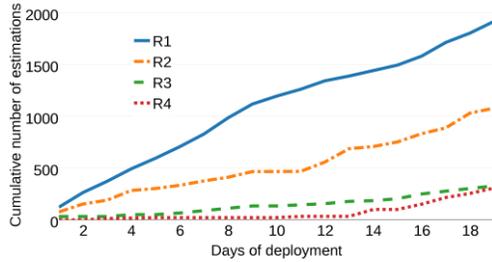


Figure 7. Cumulative number of estimations collected by each kiosk.

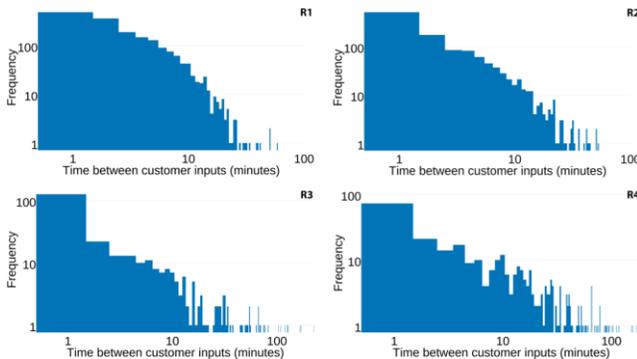


Figure 8. Distribution of time delay (minutes) between subsequent inputs on each restaurant.

To assess how many people utilized the public displays for checking queue lengths in different restaurants, we rely on quantitative data from server logs as a proxy for perceived interest in the system. The web page was loaded from 446 unique IP-addresses, and during the field trial it was accessed a total of 1796 times. Figure 9 shows the distribution of webpage hits over the 19 days of deployment and time of day. There is a significant peak on the first day of deployment. This can be mostly attributed to the novelty effect and users experimenting with the webpage, i.e. loading the page multiple times during the day to check the estimations, later stabilising on a certain user base. We note that these graphs only show webpage hits and not total amount of people that checked the estimations daily, i.e. it does not account for those that got this information from the public displays around campus. Different restaurant menus were loaded 3119 times, with R1 menu being the most popular (1117 hits), then R4 (785), R2 (667) and R3 (545).

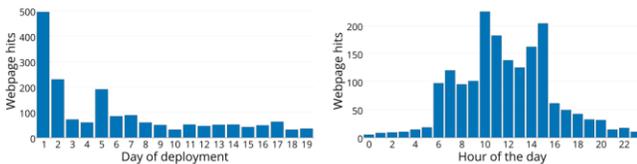


Figure 9. Distribution of webpage hits over the 19 days of deployment (left) and time of day (right).

Next, we assess the accuracy of our system’s prediction. This required us to collect ground truth manually, since no restaurant collected this information. In total we collected 381 ground truth measurements, of which 189 were

collected on day 13 and 192 on day 17 of our deployment. The measurements were conducted in the most popular restaurant (R1), from 10am till 4:30pm. In both days the queue waiting time was measured with sampling rate of two minutes.

Every two minutes we “marked” the last customer in the queue, and when the customer eventually paid we recorded the time in the original 2-minute slot.

Using the ground truth we are able to estimate the accuracy of our system’s predictions. On day 13 the kiosk was located at the back of the queue, and on day 17 the kiosk was located at the front of the queue (next to the cashier). We moved the kiosk to the front of the queue on day 17 in order to test if we could improve the estimations based on the results of Study 1 depending on the positioning of the kiosk. Since the back of the queue is a moving target, we positioned the kiosk where the flow of customers funnelled towards the queue and where said queue was fully visible. We found that the mean absolute error in our system’s prediction was 2.9 minutes on day 13, and 3.6 minutes on day 17. Figure 10 shows the real waiting time and the absolute estimation error on day 17.

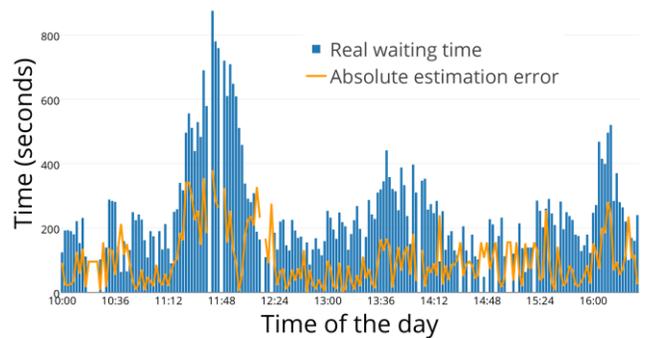


Figure 10. Real waiting time and absolute estimation error on day 17 over time of the day.

However, during our deployment we did not actively calibrate the estimation data from kiosks, and had used an arbitrary window of 15 minutes to generate our predictions. Therefore, we expect that our system should be able perform much better. In Figure 11 we show the accuracy of our system under varying window sizes (Equation 1), and after calibrating according to the findings of Study 1. The calibration process involves either subtracting .89 minutes from each kiosk estimation from the back of the queue or adding 1.59 minutes to each kiosk estimation from the front of the queue.

This process helped us identify the optimum performance that the system can achieve. When the kiosk is positioned at the back of the queue, we can correct the crowd estimation by subtracting .89 minutes from each estimation and use a sliding window of 10 minutes (Figure 11), achieving a mean absolute error of 121.45 seconds ($p < .05$). Similarly, when the kiosk is positioned at the front of the queue we can correct the crowd estimation by adding 1.59 minutes to

each estimation and use a sliding window of 8 minutes (Figure 11), achieving a mean absolute error of 126.26 seconds ($p < .05$). Thus, our system can make predictions with an error of about 2 minutes after calibration and optimisation.

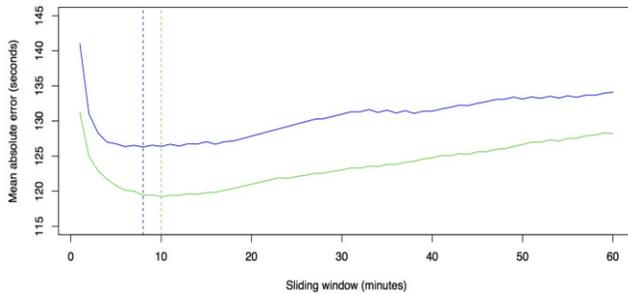


Figure 11. For varying sliding window length (x-axis) we calculate the error in our system’s prediction (y-axis). The dashed line indicates the window length that produces the smallest error. Blue: assuming kiosk is at the front of the queue. Green: assuming kiosk is at the back of the queue.

Interview and Survey Results

During Study 2 we interviewed (N=27) and collected survey answers (N=24) from customers of the participating restaurants about their overall experience with the system. We approached interviewees as they moved away from the displays after they spent a few seconds looking at the screen, while the survey was sent through internal mailing lists. We inquired about their use of the kiosks, changes to their decision making on where to go have lunch, their adopted strategies regarding the use of the system, and finally how the system could be improved. The consensus was that the concept was intriguing and useful, and something they would like to use frequently.

The majority of respondents reported having used the kiosks multiple times and were happy with the ease and effortlessness of the interaction. Others stated only being consumers of the system by checking the estimations via the public displays/webpage. Participants reported that the queue estimations were easy to comprehend. Further, participants reported that they initially looked at the graphical representation and reverted to the textual representation when two or more restaurants showed the same amount of coloured silhouettes. Participants also felt that the graphical method of representing queue size worked well with public displays allowing for quick at-glance information acquisition.

As mentioned by participants in our initial on-site interviews before Study 2, the decision on which restaurant to visit was mainly influenced by proximity. However, most participants felt that the projected waiting time does have an effect when deciding the optimal period to have lunch. Participants stated that this decision is often based on personal experience, but with the deployed system they were able to make better informed decisions and avoid having to come back later. Further, restaurant proximity

was a negotiable attribute if projected waiting time in a more distant restaurant was significantly lower. A waiting time of 5 to 7 minutes was perceived acceptable, with anything above that leading to reconsidering the destination. Ultimately, 40% of participants reported having changed, at least once, the restaurant or time they went for food.

Participants reported mostly using personal computers to check the queue situation before deciding where to go to have lunch. The public display’s queue estimations were considered a valuable addition when participants forgot to check the queue situation beforehand. Participants reported changing their mind on which restaurant to go as they were heading for lunch, if the queue in their chosen restaurant was long (over 7 minute waiting time), or simply deciding where to go on the first encountered display. One participant mentioned the possibility of adding more public displays:

“It would need more visibility in the hallways, so you can evaluate how the queues are while walking around the campus, without having to pick up the phone.”

When asked about how the system could be improved, some participants expressed hopes that the queue information could be integrated to the restaurants’ front-page.

The importance of having a large user base was highlighted by one respondent:

“More people should use it so that the information would be more accurate.”

Similarly, another respondent hypothesised about the effect of a permanent deployment on queue sizes:

“If it were in use in larger scale,[...], I am hoping it would make people notice that you can go eat earlier or later, and create a balance in the queues throughout the day.”

One respondent suggested changing the UI into a map where people can mark where the queue is right now. Based on this indication the system should be able to estimate the waiting time assuming that queue speed is consistent. Another respondent suggested a native smartphone app to request queue size and provide recommendation based on personal preferences.

When discussing about additional features, one participant suggested including a food rating functionality to help visitors decide where to have lunch. Finally, participants agreed that there was no need to add social networking features, as there are other channels to ask people when and where they want to go to have lunch.

DISCUSSION

In the New York Times article ‘Why Waiting is Torture’ [38], Richard Larson explains how “the psychology of queuing is more important than the statistics of the wait itself”. People’s expectations affect their feelings about

lines, uncertainty can magnify the stress of waiting, while feedback on the expected wait times can conversely improve the experience [22]. Obtaining feedback in a queueing situation is extremely challenging, and thus likely to compound the stress associated with waiting.

With this in mind, we set out to create a queue estimation system for collaborative environments where a single service provider is responsible for multiple service points. The system was trialled in a university environment where customers could choose amongst four separate restaurants. The main goal was to assess the feasibility and accuracy of our system in its projection of waiting time, and to establish whether this would help customers make an informed decision on where to eat. Providing wait time projections in restaurant settings can be difficult and expensive to realize using automated tools such as sensor networks or machine vision based systems [18,29]. For this reason, we wanted to investigate the feasibility of using the actual people in the queue, the crowd, as providers of waiting time estimations. We now assess our system in terms of the following criteria:

- Feasibility: Can crowdsourcing provide a method for generating reliable wait time estimations for service providers residing in close physical proximity to one another?
- User perceptions & appropriateness: Would individuals be willing to contribute to such a crowdsourcing service, and how would they perceive its usefulness?

Feasibility of Crowdsourced Queue Predictions

One of the main focus points of this paper was to determine users' willingness to input waiting time estimates. The two main issues with utilising crowd-contributed queue length estimations are: i) the frequency with which inputs are received (i.e. does intermittency play a role in generating accurate predictions), and ii) the accuracy of the estimations that people make.

Regarding the first point, we found that overall about 7% of customers are willing to input estimations. This ratio was sufficient for our system to generate predictions. However, we found that the physical location of the kiosks can substantially affect the number of user inputs. This became prominent in R4, when on day 13 (Figure 7) the kiosk was moved to a location with better visibility, increasing the number of user inputs considerably. While a higher number of inputs will naturally improve the reliability of the estimation, larger queues can also make it more difficult for customers to correctly estimate the queue size (as seen in Figure 10).

As for the second point, a key pitfall of any crowdsourcing-based system is the reliability of the contributed data [8]. The data received from customers of the restaurants are subjective and, as demonstrated in Study 1, the estimates

are biased depending on the position on the kiosk relative to the queue. Specifically, Study 1 showed that participants who provided estimations at the back of queue overestimated the waiting time, while those at the front of the queue underestimated the waiting time. Carmon & Kahneman [6] have noted that positive emotions (like finally leaving the queue) can influence people's view on their queuing experience towards a more positive perspective. Therefore, this can explain why participants in our study underestimate queue waiting time behind them after they had been served. On the contrary, those who just joined the queue may be pessimistic and therefore overestimate the waiting time. Further, we note that our presented approach is better suited for collaborative service environments, which customers visit frequently (e.g., campuses, industrial complexes). For locations in which there is a lower ratio of customers that visit frequently (e.g., large malls), it may be useful to provide users with some hints to help them make a more informed estimate (e.g., average time to serve a customer).

Crucially, the bias we identified in people's estimation can be corrected automatically, and we showed how this can lead to a substantial improvement in our system's accuracy, ultimately minimizing its error to approximately 2 minutes. This result improves on previous findings on automatic waiting time estimations using mobile phones, which have reported 2-3 minutes mean absolute error [4]. Even when considering our arbitrarily chosen 15 minute sliding window, the accuracy was only a few seconds worse in both the back and the front of the queue (Figure 11). Improvements to our system can potentially decrease this mean absolute error even further. For instance, by providing users with a suggestion or a default value based on current estimations can further lessen the initial barrier for contribution. This in turn would lead to more inputs and potentially higher overall accuracy. The estimations could also be improved by filtering out inputs that deviate significantly from the system suggestion within a certain period of time since the last input.

In summary, while improvements can be made to our system regarding the user interfaces and the calculation of the estimations, we find that crowdsourcing waiting time estimations is realistic, and relatively accurate.

Shaping User Perceptions

Norman claims that one of the major determinants of emotional unhappiness is fear of the unknown and uncertainty [27], and not knowing how long the wait time is can be stressful for students and teachers trying to get to class on time. Therefore, following guidelines from previous research [22,27], we took a slightly pessimistic view on the predictions provided to people on the web page and public displays: the queue length, represented by the silhouette figures (Figure 4), always displayed an overestimation of the waiting time instead of an underestimation (e.g., if the current estimated queue waiting

time was 10 minutes, then 4 silhouettes would appear coloured instead of 3). This is why major theme parks such as Disney typically overestimate wait times for rides: guests are pleasantly surprised when they get to rides ahead of schedule, and this positive affective response is projected to the entire queuing experience (this is known as the serial position effect [27,30]).

As a result, feedback during participant interviews showed that the system in general was regarded positively, and respondents felt that this type of system can help reduce waiting times. This is crucial since any impression (or halo effect [26]) created early in a service encounter will bias the rest of the interaction. Previous literature on perceived quality in the service sector and customer happiness clearly shows that the largest payback for effort spent in improving the interaction occurs from improving the perception of the early stages of the interaction, i.e. reducing the waiting time [22]. The effect of our system is that it becomes the first impression instead of the end of a long queue. Thus, before people arrive at the restaurant they already have certain expectations about the queue length and their options regarding visiting other restaurants.

Managing the Queuing Experience

We have considered a specific type of queuing environment, a collaborative queuing environment, which is a recurring phenomenon in large shared spaces such as hospitals, schools, industrial complexes, etc. We argue that in such settings, where individual service points are in close physical proximity, they do not compete for customers, and hence can be motivated to adopt a customer-driven queuing system that helps customers answer the question: should I wait in the line for this restaurant, or spend an extra 3 minutes walking to the closest alternative around the corner?

Our studies show that crowdsourcing can make 2 contributions. First, it can provide reliable data, and second it can reduce unoccupied time, which can be beneficial in multiple ways.

First, our results show that crowdsourcing is a feasible, accurate, and reliable way of obtaining waiting time estimations from people in the queue. Subsequently, the obtained estimations can help people make the informed decision to either visit the service point closest to them, or to try a more distant service point with a shorter projected waiting time. Our respondents confirmed that having access to approximate queue information is useful and that they would like to use this type of system frequently, however whether the projected waiting time would affect their decision on which restaurant to have lunch in varies for each individual. The decision to go to a farther restaurant would be acceptable in case of very long queues, even if it meant that the total time expended might be longer than staying in the original restaurant.

Second, our system can act as catalyst to motivate individuals to try their luck at a different establishment. While it may even take them more time to travel and get served at a different establishment, previous work has shown that in fact this could be beneficial. Specifically, previous research has shown that unoccupied waiting time feels longer than occupied waiting time [35], as was recently documented, for example, at Houston airport [38].

In a collaborative service environment, through crowdsourcing, it becomes possible to collect and visualise relevant information, and as such encourage people to get moving. Because our system removed the factor of uncertainty by showing how long the queue in the next restaurant is, customers could safely make the decision to better occupy their time by walking to another restaurant instead of waiting in the queue, doing nothing.

Limitations

We acknowledge multiple limitations with our approach. First, the system can have periods where there are no inputs during the considered sliding window, which results in the system not being able to provide estimations. However, this mostly happened in periods of low restaurant activity when the usefulness of queue estimations is also low. Second, in some cases we experienced issues with the positioning of the kiosk. Particularly for R4 it was difficult to find an optimal location which had power, WiFi coverage, and high visibility to the customers. Appropriate positioning of the kiosk is crucial to avoid low number of inputs, which in turn leads to less accurate estimates. Third, we only had access to daily sales from the till data, rather than receipt logs or more granular sales data. This meant that we could not use this data to improve the estimation calculations or test the system's resilience to noisy input. Fourth, the scale used in the kiosks was capped to 15 minutes. While this was based on observations in the restaurants prior to the deployment, we acknowledge that an adapting scale would be better suited to accommodate potential influx in number of customers. Finally, in order for customers to provide accurate estimations they need to be able to see the whole queue. This can become an issue when the queue, for example, wraps around a building.

CONCLUSION

We have presented a system that allows people waiting in a queue to contribute their subjective estimation on the queue length. This crowdsourced approach can help other potential customers make an informed decision on which service provider to visit, based on their personal preferences, on projected waiting time, distance to service provider, etc. By trialling the proposed system for 19 days in a university setting using four restaurants as case studies, we were able to show that crowdsourcing is a viable method of providing waiting time estimations. Results show that while customers' position in the queue affects their estimation on the waiting time, this bias can be corrected to provide more accurate estimates. In addition, by carefully

selecting the sliding window size during which waiting time predictions are made, accuracy can be further improved, regardless of whether the input is collected near the front or the back of the queue. Qualitative results from the field trial confirm that restaurant customers felt the system was usable, accurate, and helped create a more positive waiting experience.

In the future we will focus on understanding how such a system can influence human behaviour regarding service provider selection. We will also experiment with methods to make the system more robust against misinformation from potentially malicious crowd members, and attempt to minimize estimation error.

ACKNOWLEDGEMENTS

This work is partially funded by the Academy of Finland (Grants 276786-AWARE, 285062-iCYCLE, 286386-CPDSS, 285459-iSCIENCE), and the European Commission (Grants PCIG11-GA-2012-322138 and 645706-GRAGE). We also thank Aku Visuri, Teemu Partanen, Nemanja Vukota and Miika Keisu for their contributions.

REFERENCES

1. Didier Aubert. 1999. Passengers queue length measurement. In *Proceedings of the International Conference on Image Analysis and Processing*, 1132-1135. <http://dx.doi.org/10.1109/ICIAP.1999.797754>
2. Dietmar Bauer, Markus Ray and Stefan Seer. 2011. Simple sensors used for measuring service times and counting pedestrians. *Journal of the Transportation Research Board* 2177, 1, 74-84. <http://dx.doi.org/10.3141/2214-10>
3. D. Budic, Z. Martinovic. and D. Simunic. 2014. Cash register lines optimization system using RFID technology. In *Proceedings of Information and Communication Technology, Electronics and Microelectronics*, 459-462. <http://dx.doi.org/10.1109/MIPRO.2014.6859611>
4. Muhammed Bulut, Yavuz Selim Yilmaz, Murat Demirbas, Nilgun Ferhatosmanoglu and Hakan Ferhatosmanoglu. 2013. LineKing: Crowdsourced Line Wait-Time Estimation Using Smartphones. In *Proceedings of the 4th International Conference on Mobile Computing, Applications, and Services (MobiCase '13)*, 205-224. http://dx.doi.org/10.1007/978-3-642-36632-1_12
5. Muhammed Bulut and Murat Demirbas. 2014. Coffee shop line length monitoring using smartphones. Poster in *The 15th International Workshop on Mobile Computing Systems and Applications (HotMobile '14)*.
6. Ziv Carmon and Daniel Kahneman. 1995. The experienced utility of queuing: experience profiles and retrospective evaluations of simulated queues. Durham, NC: Fuqua School, Duke University
7. Wongun Choi, Khuram Shahid and Silvio Savarese. 2011. Learning context for collective activity recognition. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, 3273-3280. <http://dx.doi.org/10.1109/CVPR.2011.5995707>
8. Weija Dai, Ginger Jin, Jungmin Lee and Michael Luca. 2012. Optimal aggregation of consumer ratings: an application to yelp.com. NBER Working Paper No. 18567.
9. Duncan Dickson, Robert Ford and Bruce Laval. 2005. Managing Real and Virtual Waits in Hospitality and Service Organizations. *Cornell Hotel and Restaurant Administration Quarterly* 46, 1, 52-68. <http://dx.doi.org/10.1177/0010880404271560>
10. Huiji Gao, Geoffrey Barbier, Rebecca Goolsby. 2011. Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems* 3, 10-14. <http://dx.doi.org/10.1109/MIS.2011.52>
11. Jorge Goncalves, Denzil Ferreira, Simo Hosio, Yong Liu, Jakob Rogstadius, Hannu Kukka and Vassilis Kostakos. 2013. Crowdsourcing on the spot: altruistic use of public displays, feasibility, performance, and behaviours. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*, 753-762. <http://dx.doi.org/10.1145/2493432.2493481>.
12. Jorge Goncalves, Simo Hosio, Denzil Ferreira and Vassilis Kostakos. 2014. Game of Words: Tagging places through crowdsourcing on public displays. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*, 705-714. <http://dx.doi.org/10.1145/2598510.2598514>
13. Jorge Goncalves, Simo Hosio, Yong Liu and Vassilis Kostakos. 2014. Eliciting Situated Feedback: A Comparison of Paper, Web Forms and Public Displays. *Displays* 35, 1, 27-37. <http://dx.doi.org/10.1016/j.displa.2013.12.002>
14. Jorge Goncalves, Pratyush Pandab, Denzil Ferreira, Mohammad Ghahramani, Guoying Zhao and Vassilis Kostakos. 2014. Projective testing of diurnal collective emotion. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*, 487-497. <http://dx.doi.org/10.1145/2632048.2636067>
15. Donald Gross. 2008. *Fundamentals of queueing theory*. John Wiley & Sons, New York.
16. Simo Hosio, Jorge Goncalves, Vili Lehdonvirta, Denzil Ferreira and Vassilis Kostakos. 2014. Situated crowdsourcing using a market model. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*, 55-64. <http://dx.doi.org/10.1145/2642918.2647362>

17. Simo Hosio, Jorge Goncalves, Vassilis Kostakos, Jukka Riekkii. 2015. Crowdsourcing public opinion using urban pervasive technologies: Lessons from real-life experiments in Oulu. *Policy & Internet* 7, 2, 203-222. <http://dx.doi.org/10.1002/poi3.90>
18. Ching-Tang Hsieh, Hui-Chun Wang, Yeh-Kuang Wu, Liung-Chun Chang, and Tai-Ku Kuo. 2012. A Kinect-based people-flow counting system. In *Proceedings of Intelligent Signal Processing and Communications Systems (ISPACS)*, 146-150. <http://dx.doi.org/10.1109/ISPACS.2012.6473470>
19. K.C. James and M. Bhasi. 2010. Development of model categories for performance improvement studies related to airport terminal operations. *Journal of Simulation* 4, 2, 98-108. <http://dx.doi.org/10.1057/jos.2009.27>
20. Vassilis Kostakos, Tiago Camacho and Claudio Mantero. 2010. Wireless detection of end-to-end passenger trips on public transport buses. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC '10)*, 1795-1800. <http://dx.doi.org/10.1109/ITSC.2010.5625062>
21. Richard Larson. 1987. OR Forum - Perspectives on Queues: Social Justice and the Psychology of Queueing. *Operations Research* 35, 6, 895-905. <http://dx.doi.org/10.1287/opre.35.6.895>
22. David H. Maister. 1984. *The psychology of waiting lines*. Harvard Business School.
23. Leon Mann. 1969. Queue culture: The waiting line as a social system. *American Journal of Sociology*, 340-354.
24. David McDonald Jr. and Robert Stammer Jr. 2001. Contribution to the development of guidelines for toll plaza design. *Journal of Transportation Engineering* 127, 3 (2001), 215-222. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2001\)127:3\(215\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2001)127:3(215))
25. Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking Questions of Targeted Strangers on Social Networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*, 999-1002. <http://dx.doi.org/10.1145/2145204.2145352>
26. Richard E. Nisbett and Timothy Wilson. 1977. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology* 35, 4, 250. <http://psycnet.apa.org/doi/10.1037/0022-3514.35.4.250>
27. Donald Norman. 2008. *The psychology of waiting lines*. Excerpt from Chapter 4, Clerks and Waiting Lines.
28. Eamonn O'Neill, Vassilis Kostakos, Tim Kindberg, Ava Fatah gen. Schiek, Alan Penn, Danae Fraser and Tim Jones. 2006. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp '06)*, 315-332. http://dx.doi.org/10.1007/11853565_19
29. Marion K. Poetz and Martin Schreier. 2012. The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management* 29, 2, 245-256. <http://dx.doi.org/10.1111/j.1540-5885.2011.00893.x>
30. Earl W. Sasser, Richard P. Olsen and Daryl D. Wyckoff. 1978. *Management of service operations: Text, cases, and readings*. Allyn & Bacon.
31. Foldit. 2015. Solve Puzzles for Science | Foldit. Retrieved February 27, 2015 from <http://fold.it/portal/>.
32. Shirley Taylor. 1995. The effects of filled waiting time and service provider control over the delay on evaluations of service. *Journal of the Academy of Marketing Science* 23, 1, 38-48. <http://dx.doi.org/10.1007/BF02894610>
33. Shirley Taylor. 1994. Waiting for Service: The Relationship between Delays and Evaluations of Service. *Journal of Marketing* 58, 2, 56-69. <http://dx.doi.org/10.2307/1252269>
34. Gail Tom and Scott Lucey. 1997. A Field Study Investigating the Effect of Waiting Time on Customer Satisfaction. *The Journal of Psychology* 131, 6, 655-660. <http://dx.doi.org/10.1080/00223989709603847>
35. Harsh Verma. 2012. *Services Marketing: Text and Cases*. 2nd edition. Pearson.
36. Yan Wang, Jie Yang, Yingying Chen, Hongbo Liu, Marco Gruteser & Richard Martin. 2014. Tracking human queues using single-point signal monitoring. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '14)*, 42-54. <http://dx.doi.org/10.1145/2594368.2594382>
37. Jens Weppner and Paul Lukowicz. 2013. Bluetooth based collaborative crowd density estimation with mobile phones. In *Proceedings of the IEEE Conference on Pervasive Computing and Communications*, 193-200. <http://dx.doi.org/10.1109/PerCom.2013.6526732>
38. NY Times. 2012. Why Waiting Is Torture. Retrieved February 27, 2015 from <http://www.nytimes.com/2012/08/19/opinion/sunday/why-waiting-in-line-is-torture.html>
39. Bolei Zhou, Xiaou Tang, Hepeng Zhang and Xiaogang Wang. 2014. Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 8, 1586-1599. <http://dx.doi.org/10.1109/TPAMI.2014.2300484>