Final version to IEEE Pervasive Computing Magazine

ConvenienceProbe: A Phone-based Data Collection and Access System for Retail Trade Area Analysis

Chuang-Wen You[1], Hsin-Liu (Cindy) Kao[2], Bo-Jhang Ho[2], Yu-Han Chen[2], Wei-Fehng Wang[3], Lien-Ti Bei[3], Hao-Hua Chu[2], Ming-Syan Chen[1]
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan[1]
Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan[2]
Department of Business Administration, National ChengChi University, Taipei, Taiwan[3]
cwyou@citi.sinica.edu.tw, {b95701241, b96118, yuhan}@csie.ntu.edu.tw, 92307036@nccu.edu.tw, hchu@csie.ntu.edu.tw, lienti@nccu.edu.tw, mschen@citi.sinica.edu.tw

***Abstract*—Systematically and quantitatively determining patterns in consumer flow is an important question in marketing research. Identifying these patterns can facilitate understanding of where and when consumers purchase products and services at physical retail shops. Collecting data on real consumers who shop at retail stores is one of the most challenging and expensive aspects of these studies. This paper introduces a phone-based data collection system, called ConvenienceProbe, for retail trade area analysis. The proposed method specifically targets local residents shopping at neighborhood convenience stores. This study deploys and tests the system by collecting real customer flow data in neighborhood convenience stores. Results show that the consumer flow data collected from the ConvenienceProbe system is comparable to that from a traditional face-to-face interview method.**

***Index Terms*—Phone sensing, data collection and analysis, consumer behavior and marketing research.**

I. Introduction

Mobile phones have become an indispensable part of our everyday lives, as they go with us everywhere. New mobile phones are equipped with sophisticated sensing, computing, and communication capabilities. Due to the ubiquity of these phones and their sensory capabilities, it is possible to build a phone-based data collection system that observes their users everywhere and capture spatially relevant information of consumption behavior. This data collection of human behavior enables geographical and quantitative analysis of where, when, and how people conduct their everyday activities non-intrusively, i.e., without disrupting human natural behaviors.

Retail Trade Area Analysis and Marketing Research

Since different chains of CVS (abbreviation for a *convenience store*) outlets actually sell similar merchandises and satisfy similar consumer needs, marketing research about trade area plays a particularly vital role in determining their business success. Retail trade area analysis [2], which is one kind of marketing research, provides a basis for understanding, quantifying, and visualizing customer flow and movement in the area around a store. This information is critical for making business decisions such as selecting the optimal store location, identifying competing stores, and placing outdoor advertisements. Since consumer flow and direction may change due to newly opened or closed stores and changing composition of local residence, regular retail trade area analysis is necessary to track a store's current trade area and customer flow.

Opportunities for Phone-based Data Collection

Phones provide opportunities to offload the process of collecting customer flow data to local residents who own and/or carry a mobile phone and are customers of neighborhood CVS stores. Offloading data collection to consumers can significantly reduce the cost of conducting quantitative marketing studies. Phones provide opportunities to *automate* the data collection process through smart sensing, detecting, and logging of customers' CVS trips. Automating data collection makes it possible to gather consumer behavior without interrupting users' activities, and reduces the human recall errors and biases found in the traditional self-reporting [12], face-to-face interview, and surveying methods.

Traditional data collection methods for consumer behavior studies are labor-intensive [4]. In these approaches, researchers recruit human field workers to observe customers entering and leaving a store. Regular customers are asked to fill out a questionnaire about their inbound/outbound paths to the store and shopping behavior. Another common human metering method involves shadowing customers and observing their behavior over time. Since human labor does not scale, data collection is often an expensive part of consumer behavior research. This study proposes the ConvenienceProbe system to reduce the cost of collecting consumer data. The proposed system organizes phones as mobile sensors and gathers customer flow data from phones.

The contribution of this study is in the design, prototypes, and evaluation of the proposed phone-based data collection system by collecting real customer flow data to CVS outlets within an area of interest. The data includes 368 store visits from the phone-based method and 90 store visits from the traditional face-to-face method. Results show that the consumer flow data collected from the phone-based method is comparable to, and in some ways superior to, that from a traditional face-to-face interview method. Furthermore, we discuss the strengths and weaknesses in data collection between the phone-based method and the traditional face-to-face method, and provide a number of design insights on how marketing research can leverage phone-based data collection for future retail trade area analysis.

II. System Architecture and Design

The ConvenienceProbe system is based on the client-server architecture. Phones carried by participants run a client-side phone application that senses participants' consumer behavior. Phones periodically transmit consumer behavior data to a backend server through the phones' wireless networks. The following subsections explain the details of the phone application and the backend server.

Phone Application

The proposed phone application includes two modules: (1) a phone user interface and (2) automated CVS trip detection with battery-efficient sensor selection. These phone application modules are described as follows.

Figure 1 shows a simple *phone user interface* that allows participants to perform data uploading and to provide evidence for their store visits with only a few button clicks. The user interface serves two main functions: (1) take a photograph of the CVS purchase receipt as evidence of each store visit, and (2) upload CVS trip data logged in a phone to a server. Participants then matched each receipt photograph to a shopping event displayed in a list of phone-detected CVS visits (Fig. 1(b)). The user interface provides two uploading methods to transfer CVS visit data from a phone to the server through (1) a direct wireless Internet connection on a phone or (2) a USB connection to a PC, in which the PC first retrieves data from the phone and then transfers data to the server.

The system implements automated CVS trip detection inside the area of interest. The Wi-Fi sensor [3] identifies when a participant transitions between places (e.g., moves from an office to a CVS outlet) or stops in one place (e.g.,

(a) The main interface

(b) The labeling interface

Figure 1. Phone's main and labeling interfaces. The main interface (Fig. 1(a)) shows a photo receipt capturing button, a data uploading button, and an exit button. The labeling interface (Fig. 1(b)) shows a list of phone-detected store visits from which a participant matches each receipt photo to.

shopping at a CVS outlet). The phone-received Wi-Fi signature is also used to recognize a specific CVS outlet by matching the phone-received Wi-Fi signature to the Wi-Fi APs previously profiled and located nearby the CVS outlet.

To protect the location privacy of participants, the phone only uploads the CVS trip data to the server. From the phone user interface, participants clicked the uploading button to initiate any data uploading. This process required participants to give explicit permission to the phone application for each data uploading operation. This client-based approach requiring explicit user permission provides an advantage in privacy protection.

Back-end Server

The backend server includes (1) the web portal, (2) data repository, and (3) data analysis and visualization tools. These server modules are described as follows.

The web portal was created for the purpose of recruiting interested individuals who can browse information about this marketing research and its data collection process. After individuals agreed to participate in this study, the web portal asked them to complete questionnaires to determine their qualification for involvement. Participants that met the necessary qualifications were asked to fill out their contact and user profile information and to create individual accounts from which they could upload data and receive micropayments. Finally, the web portal asked them to sign a consent form agreeing to release their data. Each time a participant uploaded data to the server, the website calculated the amount of money earned and the amount of data contributed to the server's data repository.

The data repository provides a centralized storage of all CVS visit data uploaded from participants. The CVS trip data uploaded in this study includes (1) raw sensor data including timestamp, GPS coordinates, cell-ID log, and received Wi-Fi signatures, and (2) phone-detected store visits and photos of purchase receipts provided by participants. The raw sensor data are used for automated store visit detection and for reconstructing consumer path to the store (described below). Participants provide photos of their purchase receipts as evidence of actual store visits, thus avoid the situation in which participants may make false claims of CVS visits in order to earn monetary incentive.

Marketing research and retail trade area analysis retrieved data from the data repository to perform customer path reconstruction and retail trade area visualization. For the customer path reconstruction, the system reconstructs consumer path for each phone-detect CVS trip. Each consumer path has three elements. (1) An inbound path is the store-arrival path taken by a customer starting from a previous destination place (e.g., participant's home or office building) until arriving at the store. (2) In-store time is the amount of time that a customer stays at the store. (3) An outbound path is the store-departure path taken by a customer leaving the store for his/her next destination. GPS data from phones was used to plot these consumer paths and form routes. For the retail trade area visualization, Figure 3(a) shows an example of a retail trade area map computed from the Bounding Wedge-Casting method [9]. This method divides a store's surrounding area into directional wedge sectors. The store is the hub at the center of all wedge sectors. For example, the retail trade area map in Fig. 3(a) contains twelve 30-degree wedge sectors. Each wedge sector grows from the store location in the center outwards to cover locations of additional customers until the cumulative number of customers exceeds a threshold, such as 80% of customers.

III. USER STUDY

The effectiveness of the ConvenienceProbe system was tested in a user study that collected real customer flow data from three competing CVS stores situated within the same area. This section describes the design and experimental results of this comparative user study, which were guided by the following inquiries:

- What was the relative *data quality* and *quantity* collected from the ConvenienceProbe system compared to the traditional face-to-face interview?
- What was the *trade area analysis* quality obtained from the ConvenienceProbe system compared to the traditional face-to-face interview?

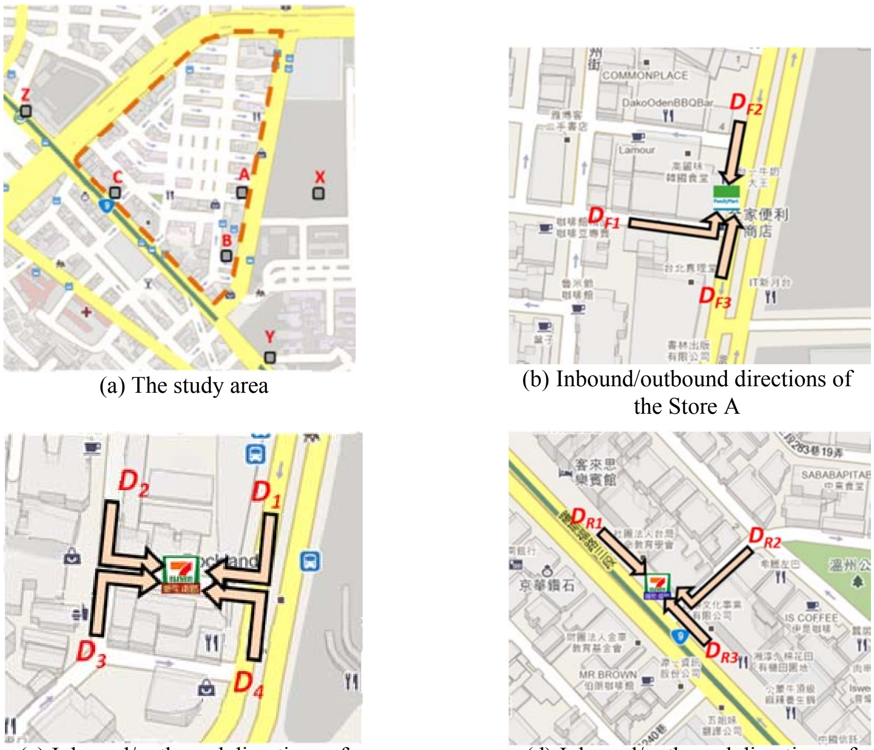
To make a meaningful comparison, this comparison study emphasizes the aspect of data collection methods, i.e., the traditional face-to-face interview method (called the *interview method* from here) vs. the ConvenienceProbe method (called the *phone method* from here), while maintaining consistency in other aspects, such as the same CVS outlets in the focal neighborhood, the same qualification criteria for selecting participants, and the same data analysis method on the customer flow (inbound/outbound paths and source/destination points).

Physical Settings

The study area was a highly-populated triangular city block near a university campus (Fig. 2(a)). This triangular area measured approximately 0.128 square kilometer, and the triangular edges were approximately 470, 540, and 640 meters long. This focal area contained three competing CVS outlets, whose locations are marked on Fig. 2(a) as stores A, B, and C.

Participants

The phone campaign recruited 42 participants who either already owned HTC Android phones capable of running our phone application or borrowed a compatible HTC Android phone from us for the duration of the study. The age distribution of participants ranged between 18 and 53, and the average age was 25. Their occupations included students, clerks, sales, engineers, housekeepers, etc. All participants were residents of, or worked in, the focal area,



(a) The study area

(b) Inbound/outbound directions of the Store A

(c) Inbound/outbound directions of the Store B

(d) Inbound/outbound directions of the Store C

Figure 2. The physical setting and inbound/outbound directions of stores A, B, and C. The triangular dotted boundary lines mark the focal area. This area is adjacent to a university campus (X). There are three CVS stores: (A) Family Mart, (B) Xinsheng 7-11, and (C) Roosevelt 7-11. Other landmarks near the focal area include Gongguan (Y) and Tai-Power (Z) metro rapid transit (MRT) stations. Store A and store C are located at 3-way T-intersections with three inbound/outbound directions, and store B is located in an alley with four inbound/outbound directions.

and regularly patronized the selected CVS stores. Compensation to participants included (1) a fixed weekly payment with a guaranteed base of approximately US$17 and (2) an extra bonus of approximately US$1 per day when the participant uploaded CVS visit data. To prevent demand artifacts, the maximum weekly payment was set to approximately US$24 (i.e., the base plus a maximal weekly bonus) so that participants would not increase frequency of their CVS visit because of the extra daily bonus. The 42 participants in this study recorded 368 visits to the three CVS stores during 3 weeks, with 167, 150, and 51 visits to stores A, B, and C, respectively.

The interview campaign surveyed 90 frequent customers, or 30 customers per store, outside of the three stores. To reduce the effect of the time-dependent factors on the profile of customers, respondents were recruited at various hours of a day (10am ~ 8pm) and both weekdays and weekends. The age distribution of the respondents ranged between 12 and 73, and the average age was 29. Their occupations were similar to the occupations of participants in the ConvenienceProbe campaign. Two-tailed T-test results show no significant differences between these two groups in gender (p=0.619), age (p=0.200), and store visit frequency (p=0.972).

Procedures of the Phone Method

The procedure consisted of two phases: (1) a screening phase to select qualified subjects and collect background information, and (2) the experiment phase. Consumers were recruited via an Internet ad and snowball sampling. The selection criteria were those who either lived or worked in the focal area, and frequently shopped at one of the three CVS stores. Qualified participants began performing data collection tasks and were asked to regularly upload their CVS visit data.

In the screening phase, interested candidates first completed questionnaires on the web portal. The questions assessed the candidates' familiarity with the digital devices, i.e., whether they had basic technical skills, and their daily consumption habits, i.e., whether they regularly patronized the selected CVS stores. After they passed the screening stage, candidates were asked to fill out a pre-study questionnaire with their personal information. They also set up accounts for uploading data later. Finally, qualified participants were asked to attend an orientation on how to use the phone-based system.

The experiment phase ran for 2-3 weeks, during which participants carried phones that recorded inbound/outbound paths of their CVS trips. Participants were responsible for recharging the phones at the end of each day.

The Interview Method

This procedure involved two steps. (1) Interviewers approached customers as they left the CVS store. (2) If the customers were regular customers of the focal CVS store, interviewers helped customers fill out a short questionnaire about the origin/destination points of their current store visit and other consumer behavior.

Evaluation metrics and results for CVS trip detection accuracy.

This study defines store detection accuracy as a metric to measure how well the CVS trip detection algorithm correctly classified events in which participants only passed by a CVS store without entering it (defined as a *non-visit*) from events where participants entered a CVS store (defined as a *visit*). The calculation of this accuracy is described as follows. First, all visit events were further classified into two types depending on whether customers made their purchases during these visits; these two types are *purchasing* and *non-purchasing* visits. To obtain the ground-truth information of these two types of visit events, we asked participants to take a photograph of the CVS purchase receipt as an evidence of each purchasing visit and keep a diary on recording non-purchasing visits. A false positive occurs when the phone detects a CVS (purchase or non-purchase) visit but in fact there is no visit (i.e., an unmatched phone-detected visit). A false negative occurs when the phone does not detect a CVS (purchase or non-purchase) visit but in fact there is a visit (i.e., an unmatched receipt photograph or dairy record). Based on the collected ground-truth information, the number of detection errors can be calculated by aggregating all false positives and negatives. The resulting detection accuracy of CVS visits was 93.75%, which was better than the detection accuracy of CVS non-visits at 90.08%. Most of the detection errors came from participants standing near the entrance to a CVS store where the phone-received Wi-Fi signature is similar to the in-store's Wi-Fi signature, thus causing detection errors. A 90% plus CVS detection accuracy rate is sufficient for consumer behavior research.

Evaluation metrics and results for inbound/outbound directions.

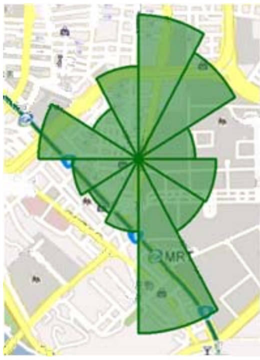
This study defines S_d as a metric for measuring similarity in customer inbound/outbound directions between the phone and interview methods. The compositions and calculations of this S_d value are described as follows. (1) All the CVS store visits were divided into different flow components based on customer inbound/outbound directions. (2) A *customer flow vector* for each CVS store was generated for both data collection methods. A customer flow vector contains vector components whose scalar values correspond to the percentages of customers entering/leaving the store from a specific direction. Figure 1 (b) and (c) show that each store has two customer flow vectors from the phone method (denoted as $\vec{V}_{phone}$) and the interview method ($\vec{V}_{interview}$). Each customer flow vector V has vector components corresponding to multiple flow directions: $\vec{V}_{interview\ or\ phone}$ = (D_1, D_2, .., D_m). For example, store A in Fig. 2(b) has three flow directions (D_1, D_2, D_3). The scalar value D_1 is the percentage of customers entering/leaving the store from/to the north direction. (3) The Cosine-based correlation between the two customer flow vectors ($\vec{V}_{phone}$ and $\vec{V}_{interview}$) was computed for each store as a measure of similarity.

$$S_d = \frac{\vec{\mathbf{V}}_{phone} \bullet \vec{\mathbf{V}}_{interview}}{|\vec{\mathbf{V}}_{phone}||\vec{\mathbf{V}}_{interview}|} \qquad (1)$$

Table 1 (Column 1 ~5) shows the similarity results for customer flow vectors of all three stores. For all three stores, similarity scores (Column 4) are higher than 92%, with an average similarity score of 96%. The similarity scores for stores A and B reach nearly 100%. For statistical validation, we conducted the Chi-square (χ^2) equality of proportion test for equality of distributions on data obtained by the phone method (i.e., the observed) and the interview method (i.e., the expected). The *p* values (Column 5) are all close to 1 and larger than the conventionally accepted significance level of 0.05. So, the null hypothesis that the two distributions are the same is not rejected for each store. In other words, the Chi-square equality of proportion test indicates a good fit between the phone method (observed) and the interview method (expected) for all three stores.

Table 1. Similarity results for the flow vectors (i.e., a store's customer inbound/outbound directions) of the phone and interview methods for all three stores and for the trade areas (i.e., the area where a store's customer come from) of the phone and interview methods for each of three CVS stores. The similarity scores (Column 5) were computed using Eq. (1). The similarity of trade areas is measured by overlapping, miss, and extra ratios (column 9, 10, and 11, respectively) defined in the subsection (Evaluation Metric and Results for Retail Trade Areas).

CVS stores Inbound & outbound directions	Flow vector from the phone method (n=42, #visits=368)	Flow vector from the interview method (n=90, #visits=90)	Similarity score (S_d)	χ^2 value	Phone method trade area (km^2) (n=42, visits=368)	Interview method trade area (km^2)(n=90, visits=90)	Intersection area (km^2)	Overlapping ratio (%)	Miss ratio (%)	Extra ratio (%)
Store A D1 D2 D3	 17.3% 27.8% 54.7%	 13.4% 36.6% 50.0%	98.42%	0.037 (*p*=0.9822)	7.631	1.959	1.695	86.5%	13.5%	77.8%
Store B D1 D2 D3 D4	 38.2% 25.7% 17.9% 17.9%	 32.2% 25.5% 25.0% 17.8%	98.56%	0.020 (*p*=0.9985)	4.520	1.071	0.963	89.9%	10.1%	78.7%
Store C D1 D2 D3	 24.1% 37.9% 37.9%	 24.3% 55.1% 20.6%	92.38%	0.199 (*p*=0.9053)	2.894	2.134	1.626	76.2%	23.8%	43.8%



(a) Store A (phone method)

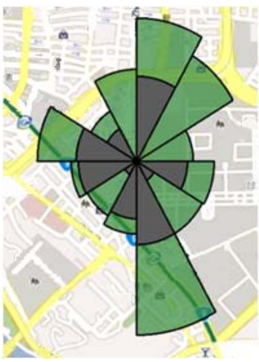
(b) Store B (phone method)

(c) Store C (phone method)

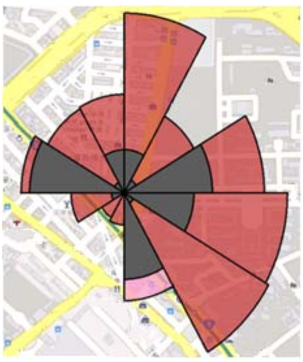
(d) Store A (interview method)

(e) Store B (interview method)

(f) Store C (interview method)



(g) Store A (overlapping phone /interview)



(h) Store B (overlapping phone /interview)

(i) Store C (overlapping phone /interview)

Figure 3. The trade areas of stores A, B and C obtained from the phone method (the upper graphs) or interviewing method (the middle graphs). The bottom row shows overlapping trade areas of these two methods. The black regions indicate the intersection regions between two methods.

Evaluation Metric and Results for Retail Trade Areas

This study defines three metrics to measure the similarity of trade areas obtained from the phone and interview methods. (1) Denote *overlapping ratio* as the overlapped trade area between the phone and interview methods divided by the trade area of the interview method (as the baseline area). (2) Denote *miss ratio* as the trade area found in the interview method, but not in the phone method, divided by the trade area of the interview method. (3) Denote *extra ratio* as the trade area found in the phone method, but not in the interview method, divided by the trade area of the phone method.

Table 1 (Column 6 ~ 11) shows the similarity results indicated by the retail trade areas of the three CVS stores. The overlapping ratio (Column 9) is high at 86.5%, suggesting that the phone method was able to capture more trade areas than the interview method. Similarly, the miss ratio (Column 10) is low at 13.5%, suggesting that the phone method missed fewer trade areas than the interview method.

The extra ratio (Column 11) is high at 77.8%, suggesting that the phone method captures a lot more trade area than the interview method. The high extra ratio in Fig. 3 (conservative retail trade areas of stores A, and B in the 30-degree wedge presentation) shows that the trade areas of stores A, B and C obtained from the phone method (upper two graphs in Fig. 3) cover the trade areas obtained from the interview method (middle two graphs in Fig. 3). The lower two graphs in Fig. 3 show the overlapping trade areas of these two methods. The reason for the high extra ratio was due to the limitations of the interviewing method (described in greater detail in the next section).

IV. DISCUSSION

This section discusses lessons learned and pros and cons associated with the phone and interview methods.

Causes of Imprecise Data: Phone Sensory Errors vs. Human Recall/Communication Errors

The phone method collected more *elaborated* and *precise* consumer flow data than the interview method. For marketing research, detailed and accurate customer source and destination points and path information are critical to obtaining accurate trade area results and for understanding the trade areas in the focal region. The reasons for the different data quality between these two methods are described as follows.

The interview method relies on human memory and human spatial cognitive capabilities to recall and communicate source and destination points during face-to-face interviews. Since face-to-face interviews were limited to only a few minutes, respondents often used colloquially-labeled landmarks as their source/destination locations, e.g., “nearby the McDonald restaurant,” “nearby the church,” etc., which were verbally quick to communicate but lacked fine-grained spatial accuracy. Additionally, respondents often had difficulty communicating their store arrival and departure paths due to a lack of spatial orientation and directions. In contrast, the phone method recorded and mapped customer flow data using sensors and software tools. Although the phones’ GPS sensors in this study had an average positional error of 10-15 meters, their sensory errors were considerably less than imprecise human communication.

Compensation Models: Paying Customers vs. Paying Interviewers

The phone method places the burden of data collection on customers who owned, maintained, and operated the data collection devices, i.e., their phones. In contrast, the interview method places the burden of data collection on interviewers who traveled to the stores and manually collected data from the customers. These two methods produced different compensation models. In the phone method, most compensation went to paying the customers for their device management time and effort. In the interview method, most compensation went to paying the interviewers for their time and effort.

Unlike the interview method, the phone method offers additional flexibility and potential for extending its compensation model to include real-time and context-based rewards such as in-store coupons, which could be delivered over each participant's phone and tailored specifically toward his/her location and historical consumer behavior. This context-relevant coupon service could increase the incentive for participation in the study and reduce the direct payouts needed. This cost saving could be significant if the study involved a large number of participants. In other words, flexibility in the compensation model in the phone method offers scalability potential for a large number of everyday users as participants.

Scalability potentials in data collection (longitudinal data)

The phone method has several scalability potentials, including longitudinal scalability, geographical scalability, and across many types of stores. These scalability potentials are difficult and/or expensive to achieve using the traditional interview methods. To marketers, the beauty of ConvenienceProbe is in connecting the phones (i.e., the phones' owner profiles) to their routes and consumption records, thus the accumulated behavioral data become more reliable and accurate than the traditional survey method. On geographical and store scalability, the phone method collects consumer flow data not just on CVS stores within a focal area, but also other types of stores, restaurants, movie theaters, etc., both inside and outside of the focal area. If retailers outside of the focal area are interested in obtaining these consumer flow data, e.g., for the purpose of mapping their own stores' trade areas, they can participate by sharing the cost of this data collection or by offering phone-based coupons as described in the previous paragraph. Aggregating compensation from multiple retailers over a large geographical area enables spatial scalability and store multiplicity. In contrast, since the interview method is only able to target customers of specific store(s), its consumer data is usually applicable to one or a few retailer(s) only, thus more difficult to scale its compensation structure to other retailers.

V. RELATED WORK

Traditional Data Collection Techniques in Marketing and Consumer Research

Traditional data collection techniques for consumer flow and trade area studies include human shadowing and recording [5], or consumer surveys [11]. Human shadowing and recording techniques require human observers at stores, who directly follow or indirectly survey ingoing/outgoing customers to determine their trajectories upon entering/leaving and/or their origins/destinations. Human observers also track the number of incoming/outgoing customers to determine the number of store visits. Other human recording [10] or surveying techniques record subtle consumer behaviors, but require intensive human labors.

Other techniques for analyzing trade area are based on the concepts of spatial monopoly and market penetration. The concentric rings methods, drive time/distance polygons or Thiessen (Voronoi) polygons, and probabilistic trade area surfaces [1] are examples of the spatial monopoly approach. These techniques are often simplified and fail to incorporate the effects of competing nearby stores. The Huff model [2] is an example of the market penetration approach, which includes the effect of competing stores by modeling a spatial variation in the proportion of households served by each surrounding store. However, this technique requires a calibration phase in which results of customer survey compare the relative attractiveness of nearby stores. Furthermore, this technique only estimates the extents of trade area without the intermediate inbound/outbound trajectories.

This study does not suggest that the ConvenienceProbe is superior to traditional consumer survey method. A traditional consumer survey can probe consumers' attitudes and motivations, while ConvenienceProbe can only record true behaviors and movements. However, in the case of collecting behavior patterns for trade area analysis, as in the current study, using consumers' own mobile phones can overcome the problems of imprecise data collection due to human memory and perception.

Mobile and UbiComp Technologies in Marketing Research

Previous studies focused on how to augment shopping experiences or promote the use of UbiComp technologies at stores. Shopping Tracker [13] is a phone-based shopping tracking system that monitors customers' in-store shopping time. This system uses a phone's movement sensor to recognize unique shopping movement trajectories resulting from store aisle layouts. Moiseeva *et al.* [8] detected the activities of retail shoppers, including their transportation modes and in/out building activities, from their multi-day GPS traces. They investigated the potential benefit of incorporating GPS, cell phone, and RFID technologies to reduce the self-reporting efforts of respondents regarding their shopping trips. The data collected by ConvenienceProbe is similar to [8]; however, ConvenienceProbe goes a step further to analyze the data for mapping retail trade areas. May *et al.* [7] collected GPS trajectories and developed a model to infer the population of people passing by outdoor poster campaigns at different time and day. Liebig *et al.* [6] further focused on the population of people passing by indoor train station poster campaigns. Since GPS signals are not available indoors, their system measured the passage frequency of the floors and stairways as data for probability inference.

VI. CONCLUSION & FUTURE WORK

This study presents a novel application of phone-based data collection to consumer behavior and marketing research. Mobile phones were used to outsource consumer data collection to everyday consumers, and to automatically detect the behavior patterns from their phones. This phone-based data collection system was deployed to collect real customer flow data from neighborhood convenience stores. A comparison user study shows that it is possible to use mobile phones to collect quality consumer flow data and obtain accurate spatial patterns. The results of this approach are comparable to those from the traditional face-to-face interview method. Given the ubiquity of mobile phones, this study opens a door to the practical use of phones from everyday consumers to sense and report consumer behavior in marketing research. In other words, this study goes beyond traditional *consumer self-reporting* to UbiComp's *phone automated-reporting*, i.e., toward an invisible data collection method that does not affect natural consumer behavior.

The current ConvenienceProbe system has several limitations, which are also directions for future research. First, the current data collection software only runs on several selected HTC Android mobile phones, which limits participation to those who have these HTC phones. Future work should extend this data collection software to other major phone platforms. Second, the current GPS sensor errors create problems in customer path/trajectory reconstruction. Future work should incorporate advanced post-processing techniques, such as pedestrian walkways constraints, to improve the accuracy of customer path/trajectory reconstruction.

REFERENCES

[1] B. Gross, "Dynamic Trade Areas. Business Geographics," pp. 30-33, September (1997).

[2] D.L. Huff, "A Probabilistic Analysis of Shopping Center Trade Areas," Land Economics, vol. 39, no. 1, pp. 81-90, 1963.

[3] D.H. Kim, Y. Kim, D. Estrin, and M.B. Srivastava, "SensLoc: Sensing Everyday Places and Paths using Less Energy," Proc. ACM Conf. Embedded Networked Sensor Systems (Sensys '10), pp. 43-56, 2010.

[4] R.J. Kuo, S.C. Chi, and S.S. Kao, "A Decision Support System for Selecting Convenience Store Location Through Integration of Fuzzy AHP and Artificial Neural Network," Computers in Industry, vol. 47, no. 2, pp. 199-214, 2002

[5] M. Levy, and B.A. Weitz, Retailing Management, 7th Edition. McGraw-Hill Higher Education, 2009.

[6] T. Liebig, H. Stange, D. Hecker, M. May, C. Körner, and U. Hofmann, "A General Pedestrian Movement Model for the Evaluation of Mixed Indoor-Outdoor Poster Campaigns," In Proc. Int'l workshop on Pervasive Advertising, 2010.

[7] M. May, C. Körner, D. Hecker, M. Pasquier, U. Hofmann, and F. Mende, "Handling missing values in GPS surveys using survival analysis: a GPS case study of outdoor advertising," Proc. Int'l Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD '09), pp. 78-84, 2009.
[8] A. Moiseeva and H.J.P. Timmermans, "Imputing Relevant Information from Multi-day GPS Tracers for Retail Planning and Management using Data Fusion and Context-sensitive Learning," J. Retailing and Consumer Services, vol. 17, no. 3, pp. 189-199, 2010.
[9] A.R. Patel, T.J. Fik, and G.I. Thrall, "Direction Sensitive Wedge-casting for Trade Area Delineation," J. Real Estate Portfolio Management, vol. 14, no. 2, pp. 125-140, 2008.
[10] "Retailers Improve Service with Smart Security," http://www.asmag.com/showpost/9586.asp.
[11] T.J. Stanley, and M.A. Sewall, "Image Inputs to a Probabilistic Model: Predicting Retail Potential," J. Marketing, vol. 40, no. 3, pp. 48-53, 1976.
[12] M. R. Solomon. Consumer Behavior, 9th ED, pp 143-145, 2011.
[13] C.-W. You, C.-C. Wei, Y.-L. Chen, H.-H. Chu, and M.-S. Chen, "Shopping Time Monitoring in Physical Stores Using Mobile Phones," IEEE Pervasive Computing, vol. 10, no. 2, pp. 37-43, 2011.
[14] C.-W. You, C.-C. Wei, Y.-H. Chen, J.-C. Hu, W.-F. Wang, H.-H. Chu, L.-T. Bei, and M.-S. Chen, "Convenience Probe: A Participatory Sensing Tool to Collect Large-scale Consumer Flow Behaviors," ACM UbiComp 2010 Poster.