On Hyper-local Conversational Agents in Urban Settings

Utku Günay Acer
utku_gunay.acer@nokia-bell-labs.com
Nokia Bell Labs, Antwerp

Marc van den Broeck
marc.van_den_broeck@nokia-bell-labs.com
Nokia Bell Labs, Antwerp

Fahim Kawsar
fahim.kawsar@nokia-bell-labs.com
Nokia Bell Labs, Cambridge

ABSTRACT

Conversational agents are increasingly becoming digital partners in our everyday computational experiences. Although rich, and fresh in content, these agents are completely oblivious to users' locality beyond geospatial weather and traffic conditions. In this position statement, we envisage a brand-new class of conversational agents that are hyper-local, embedded deeply in a local neighbourhood, e.g., at urban landmarks - providing rich, purposeful, detail, and in some cases playful information relevant to a neighbourhood. By design, these agents are spatially constrained, and one can only interact with them once she is in close vicinity at street-level granularity. Learning from quantitative (n = 1992) and qualitative (n = 21) studies, we identify a set of information that these agents must accommodate. Finally, we discuss the technical architecture of this class of conversational agents that leverages covert communication channel, edge AI and on-body devices for offering such hyper-local information access.

CCS CONCEPTS

• **Human-centered computing** → **Sound-based input / output**; **Ubiquitous and mobile computing systems and tools**; *Collaborative content creation*; *Mobile computing*;

KEYWORDS

Conversational Agent, Citizen Engagement, Edge AI, Spontaneous Interaction

ACM Reference format:
Utku Günay Acer, Marc van den Broeck, and Fahim Kawsar. 2019. On Hyper-local Conversational Agents in Urban Settings. In *Proceedings of The 5th ACM Workshop on Mobile Systems for Computational Social Science, Seoul, Republic of Korea, June 21, 2019 (MCSS'19),* 4 pages.
https://doi.org/10.1145/3325426.3329949

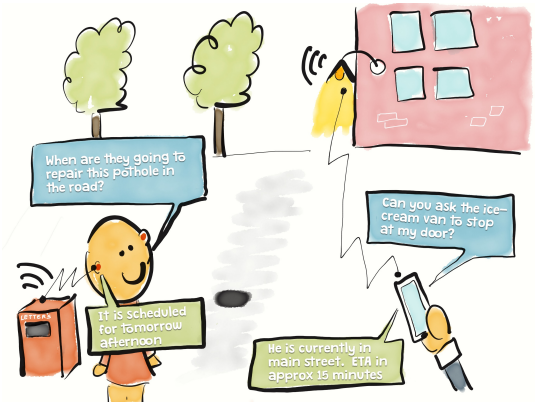

Figure 1: Hyper-local conversational agents are embedded deeply in local landmarks and offer rich, purposeful, detail information relevant to a neighbourhood in a spatiotemporal manner.

1 INTRODUCTION

Location-based services have been a widely investigated field in mobile computing. Typically, once a user invokes a service request through a mobile device, the device acquires the current location and retrieves a service that is dependent on the user's spatial and temporal context from a remote server [6, 10]. We have seen a remarkable surge of this class of applications offering a variety of experiences including navigation support, recommendation for venues, augmentation of a search for people, places, things, etc. Although useful, these applications often lack a locality view both temporally and spatially, i.e., information that is only available locally to a local citizen. For instance, consider what if one would like to know the subjective aspects on a neighbourhood street, such as how crowded, loud, or clean it is at different hours of the day, or the timing of certain local events - postman passing by, an increase of pollen level etc. This extreme local information, today, unfortunately, is not available at global scale. Sensory and crowdsourcing systems are now being developed to accommodate such fine-grained view of a neighbourhood including both quantitative and qualitative fronts [1, 2, 4, 5, 7, 8].

In this work, we envision a brand-new system that offers fine-grained local information of a neighbourhood to its citizen. Taking a user-centred view, we report two studies that at first quantitatively look at the information affinity - different information citizens expect in an urban environment beyond what is available today and then extrapolate the modality of interaction for such information access qualitatively. Based on our findings, we propose the development of a hyper-local conversational agent that is embedded deeply in a local urban landmark and offers vibrant, purposeful, detail, and in some cases playful information relevant to a neighbourhood. Figure 1 illustrates the primary functional behaviour of this proposed system conceptually. While the interaction with these agents can be done with the textual form, we have identified

MCSS'19, June 21, 2019, Seoul, Republic of Korea

ACM ISBN 978-1-4503-6777-6/19/06…$15.00
https://doi.org/10.1145/3325426.3329949

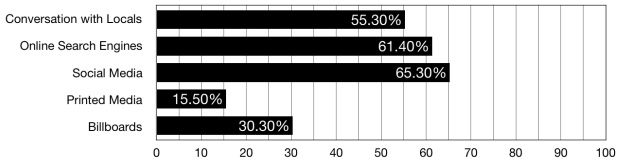

Figure 2: Preferred methods for accessing local information

a strong polarity towards spoken conversations from our participants. Besides, reflecting on our study findings, we determined that users prefer spontaneous and app-free access to such information. We have accommodated these design guidelines in our system and developed the main components of our systems borrowing principles from covert communication, semi-stateful processing, and flexible conversation management.

In what follows, we first report the two studies, and then we present the initial design of our hyper-local conversational agent.

2 CONTEXTUAL STUDIES

We begin by reporting two studies. The first study aims at uncovering the variety of spatiotemporal information that a citizen wants to have from her neighbourhood. The objective of the second study is to understand this degree of information qualitatively, and extrapolating on the modality of the information accessibility.

2.1 Online Survey

The purpose of this survey is to understand the type of information users prefer to acquire concerning their neighbourhood in a spatiotemporal manner. We are mainly interested in determining 1) a set of information types and 2) the frequency of interaction, and 3) the preferred modality of communication.

We have used Amazon Mechanical Turk[1] that allowed us to collect responses from 1992 participants in 5 continents (51.9% male, 48.1% female, age range 19-70 with 58.1% between 25 and 40). Within a range between 1 and 10 (1 lowest, 10 highest), 87% of the participants ranked their digital mindset above 5.

Figure 2 shows the preferred medium for the participants to access local information. In addition to conversations with locals (55.3%), they typically use social media (65.3%), search engines (61.4%), billboards (30.3%) and city magazines (15.5%) to access local information. 90% of the participants indicate that they would like to be more informed about their surroundings in the city they live. While 81.4% said they can rather easily access local information, 88.9% of the people said they would be interested in more effective approaches that do not require heavy filtering.

On *Information Affinity*, Figure 3 presents the types of information about their locality that the users would like to acquire. 45% of the users indicated they would make health-related queries to gather information including air quality, allergens, noise level, street cleanliness, etc. For 50.9%, community-related questions to provide information about new shops, social events, places with the best overall mood, crowded areas, etc. 59.2% of the people have said that they are like to inquire infrastructure related content such as public transportation, planned street works in an area. Safety related requests such as contacting the police, receiving and submitting warnings are important for 48% of the people. 54.3%, on the other hand, is interested in questions regarding the city including places to buy a certain product, places to go for certain activities, etc.

On *Interaction Frequency*, 41.4% of the participants preferred to access this locality information once a week whereas 27.2% multiple times a week, 10% once a day, and 4.1% multiple times a day respectively.

On *Interaction Modality*, we notice that most people mentioned that they use a conversational agent in public. It is the preferred medium for 24.3% of the participants. 40.5% of participants say that they would use it when they are alone whereas 13.9% remarks that they avoid using such an agent due to the noise in the city.

2.2 Semi-Structured Interviews

The survey offered us a quantitative view on information affinity, access frequency, and interaction modality. We wanted to understand the reasons behind the trends we reported earlier. To this end, we interviewed 21 individuals (13 men, 8 women, age range 25 - 72) following an interview technique called laddering[2] to uncover the core reasons behind users responses. We analysed the interview data by coding the individual responses using affinity diagram to derive final themes.

On *Information Affinity*, the participants essentially echoed our survey responses highlighting their desire for neighbourhood information concerning health outbreaks (e.g., pollen, flu, etc.), community events (e.g., street party, local school events, etc.), infrastructure related (e.g., construction schedule, noise management, etc.) and safety related (e.g., petty crime, etc.) Most participants have acknowledged that this system can help them engage with their neighbourhoods in a more informed manner. Some participants ($n = 15$) mentioned that they are not as aware of the city they live as they would have desired. They typically use social media or monthly city magazines to learn about events where they live in; however, filtering information generally is a problem.

On *Access Frequency*, majority of the participants mentioned while they would prefer to have such information capacity at the tip of their fingers, they won't access it daily unless there is an urgency, e.g., safety or infrastructure-related information. However, on detailing further, we have identified that weekly access has the most substantial polarity, which matches with our quantitative finding.

On *Interaction Modality*, 8 participants have expressed that they prefer voice-based interaction for such information access. Multiple of them mentioned about conversational agents commercially available today, e.g., Siri, Alexa, etc. as an example for interaction medium. Three of them have said they would have used the agent provided with their devices if it became more socially acceptable.

When asked about what they think of audio as a communication medium, Sixteen people indicated that they like audio as a user interface if it is presented discreetly in a public setting, e.g., over a headphone. However, It was pointed out that the audio feedback

[1] https://www.mturk.com/

[2] http://www.uxmatters.com/mt/archives/2009/07/laddering-a-research-interview-technique-for-uncovering-core-values.php

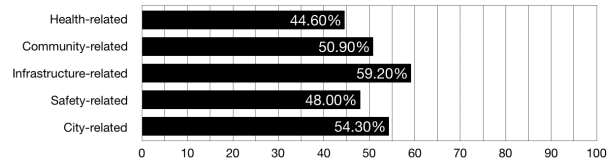

Figure 3: Types of Preferred Hyper-local queries

needs to be short and it needs to arrive at the right time to avoid any disruptions in user activity.

Overall, all participants have expressed that they would have such a system. One concern people have is that it can be abused by a provider to flood users with advertisements or it can lead to information overload. As a remedy, some users noted that the system must not just push information, but it should only respond to specific queries. Another point of discussion was privacy. While people are willing to share limited data to receive, they are most interested in inquiries of public information such as availability in an enterprise, any event in the area, housing, health conditions, public transport, etc. Because the system provides spatiotemporal public information rather than a personal service that requires user data, it perceived positively.

Based on the findings of these two studies, we have designed our hyper-local conversation system for an urban environment that we present in the next section.

3 SYSTEM OVERVIEW

In this section, we provide a descriptive view of the proposed hyper-local conversational system. The architecture of this system is depicted in Figure 4 that is composed of a collection of agents. We envision the agents run embedded to local landmarks in an urban landscape such as a light post, a building, a tower, each responsible for its vicinity to provide spatio-temporally relevant information. The system is composed of several technical building blocks including the following:

- Communication Unit: These agents have local connectivity through WiFi. They either run on WiFi Access Points (AP) that also has computation and storage capability or on other devices that are directly connected to APs. Through covert WiFi communication channels [3], user devices discover the agents and initiate conversational protocols.
- Interaction Unit: The agent uses raw audio signal from the user devices requesting service. Locally running, embedded machine learning mechanisms then interpret these signals to understand user requests. These requests are then matched with the knowledge base composed from various local data sources. The responses to user requests are sent back to user devices in the form of text that is then converted to the audio to play it back to the user.
- Knowledge Base: The agents aggregate information from local data sources and use APIs from nearby resources to create a knowledge-base and let people access it. Sensory systems driven by urban IoT [1, 7] or crowd-sourced systems

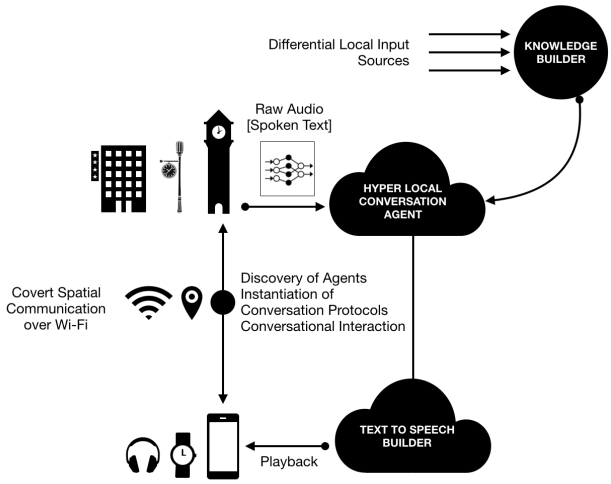

Figure 4: The architecture of the hyper-local conversational system.

such as the one reported in [2, 9] can be effectively used to build such knowledge base about a locality.

While the hyper-local conversational agents may use resources provided by edge or cloud computing for computationally heavy tasks, their scope is limited to only their vicinity and users nearby. We consider remote requests are handled in servers through traditional methods hosted on the global Internet. With an agent running on a large number of landmarks of the architectural landscape of an urban setting, we can present finely-grained spatiotemporal service to interested users. This service is not limited by a user device or the set of applications a user installed in their device, and the user data. Instead, a user can request any information that is relevant to its current location.

The proposed solution is by design constrained spatially and offers spontaneous and stateful interactions. As a user moves from one point to another that are served by different agents, the computation and data that is associated with serving the request move through these agents.

As an example scenario: consider a user asking the conversational agent if a café that is nearby, serves a particular blend of coffee and has empty seats at the moment. The agent checks the menu of every close-by establishment to see which ones serve the desired coffee blend. Then, it uses the APIs from the corresponding places to inquire whether they have available seats. Based on the results, the agent then makes a suggestion to the user and proposes to the user to give turn-by-turn directions to the suggested café. If the user complies, the agent then gives the first direction. As the user moves towards the destination, their device will opportunistically connect with another agent instance. We provide necessary mechanisms to ensure that this instance is informed about the user receiving directions for a particular destination. Then, this instance will provide another direction for the user to follow until the destination.

4 DISCUSSION AND OUTLOOK

In this position paper, we lay out our vision for a hyper-local conversation agent. This agent runs locally on urban landmarks and aggregates information from local sources to create a knowledge-base that contains every detail about their surroundings. The user devices opportunistically connect to these agents leveraging the wide coverage of WiFi and retrieve the relevant local content using spoken commands. The service the users receive is spontaneous, flexible and stateful requiring user state to move across multiple agents reflecting the users' mobility. Unlike state-of-the-art location-based services and traditional virtual assistants, this agent relies mostly on public information. Its scope is not limited by the applications users install on their devices.

We have conducted two user studies to understand whether such a system is reasonable for potential users. In either case, the participants have noted they would have used such a system to better engage with the urban setting they inhabit if some concerns are addressed. The current methods they use require various applications that depend on the content of the request and/or manual filtering of query results.

Such a system needs to address several issues. For example, a service from the agent should only be invoked by users to prevent unwanted advertisements. In addition, mechanisms that block information overload are necessary. Using a conversational agent in public is a social issue that is perceived to be not widely acceptable. As mobile devices further transform daily lives, this will arguably no longer be a problem for adoption.

ACKNOWLEDGMENTS

This work is part of the ESSENCE project. ESSENCE is an icon project realized in collaboration with imec and with project support from VLAIO (Flanders Innovation & Entrepreneurship).

REFERENCES

[1] imec - city of things. https://www.imec-int.com/en/cityofthings. Accessed: 2019-05-08.

[2] U. G. Acer, M. van den Broeck, C. Forlivesi, F. Heller, and F. Kawsar. Scaling Crowdsourcing with Mobile Workforce : A Case Study with Belgian Postal Service. In *Proceedings of the 2019 ACM International Joint Conference and 2019 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '19, 2019.

[3] U. G. Acer and O. Waltari. WiPush: Opportunistic Notifications over WiFi Without Association. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MobiQuitous 2017, pages 353–362, New York, NY, USA, 2017. ACM.

[4] F. Alt, A. Sahami Shirazi, A. Schmidt, U. Kramer, and Z. Nawaz. Location-based crowdsourcing: Extending crowdsourcing to the real world. pages 13–22, 01 2010.

[5] G. Chatzimilioudis, A. Konstantinidis, C. Laoudias, and D. Zeinalipour-Yazti. Crowdsourcing with Smartphones. *IEEE Internet Computing*, 16(5):36–44, Sept 2012.

[6] S. Dhar and U. Varshney. Challenges and Business Models for Mobile Location-based Services and Advertising. *Commun. ACM*, 54(5):121–128, May 2011.

[7] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan. The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, MobiSys '08, pages 29–39, New York, NY, USA, 2008. ACM.

[8] D. Hristova, A. Mashhadi, G. Quattrone, and L. Capra. Mapping Community Engagement with Urban Crowd-Sourcing. In *Proc. When the City Meets the Citizen Workshop (WCMCW)*, pages 14–19, Palo Alto, CA, USA, June 2012. AAAI.

[9] S. S. Kanhere. Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces. In C. Hota and P. K. Srimani, editors, *Distributed Computing and Internet Technology*, pages 19–26, Berlin, Heidelberg, 2013. Springer.

[10] A. Kupper. *Location-based Services: Fundamentals and Operation*. John Wiley & Sons, Inc., USA, 2005.