Smartphones

Editor: Roy Want ■ Google ■ roywant@acm.org

From Smart to Cognitive Phones

Andrew Campbell and Tanzeem Choudhury

EDITOR'S INTRO

Over the years, smartphones that integrate sensors have enabled many novel mobile capabilities. This article examines the trend in sensor-based applications, and where it might ultimately lead us. *—Roy Want*

As smartphones get smarter by utilizing new intelligence in the phone and the cloud, they'll start to understand our life patterns, reason about our health and well-being, help us navigate through our day, and intervene on our behalf. Here, we present various smartphone sensing systems that we've built over the years, arguing that, eventually, smartphones will become cognitive. First, however, we look back at how sensing capabilities in phones evolved.

CLOSE TO THE ARCTIC CIRCLE

In February 2005, Nokia brought together two dozen researchers from academia and the Nokia Research Center to brainstorm about the future of sensor networks. The workshop, held close to the arctic circle in Kuusamo Finland, was organized by Henry Tirri (now CTO of Nokia) around the theme of large-scale sensor networks.

There was considerable discussion about how the phone could serve as a user interface and gateway for existing sensor networks. However, for many attendees, the workshop reoriented their view of future sensor networks. The view moved away from traditional embedded sensor devices and applications, best typified by motes and environmental monitoring, and more toward using the phone as a sensor in various applications, such as those related to healthcare, traffic monitoring, or gaming. The workshop led to the creation of the Nokia SensorPlanet Project (www.research.nokia.com/research/projects/sensorplanet).

It wasn't until early 2007 that Nokia released a phone with an embedded accelerometer. Interestingly enough, Nokia didn't mention the accelerometer in that N95 phone or provide an API to access the data, because the accelerometer was only there for video stabilization and photo orientation. However, when Nokia's Péter Boda (the SensorPlanet leader) visited Dartmouth College, he casually mentioned that the N95 included an embedded accelerometer—later that year, the API was publicly released. This, along with the phone's GPS sensors, led researchers at Dartmouth and elsewhere to study and develop new phone-based sensing applications, such as the CenceMe app.[1]

CenceMe implements classifiers directly on the phone to infer the user's physical activity (sitting, walking, or running) and social interaction (whether or not the user is having a conservation, for example) in real time. Then it shares this sensing presence with the user's social network friends on Facebook. CenceMe, originally implemented on the N95 in 2007, was ported to the iPhone and released when the App Store first opened in 2008.

SMARTPHONE SENSING

If we fast forward to today, the rate of change in mobile phones has been staggering. Today's top-end smartphones come with 1.4-GHz quad-core processors and a growing set of inexpensive yet powerful embedded sensors. These smartphones include an accelerometer, a digital compass, a gyroscope, a GPS, quad microphones, dual cameras, near-field communication, a barometer, and light, proximity, and temperature sensors. They also have multiple radios for body, local, and wide area communications; 64 Gbytes of storage; and the touchscreen. Furthermore, application delivery channels, such as the App Store, are transforming phones into app phones capable of downloading myriad applications in an instant.

Each new smartphone release offers advances in sensing, computation, and communications—Moore's law in its new multicore form keeps marching on. The OS wars of the desktop era are being revisited, with Apple's iOS and Google's Android leading the charge. New breakthroughs in HCI are changing how we interact with phones—just consider Apple's Siri, which uses voice

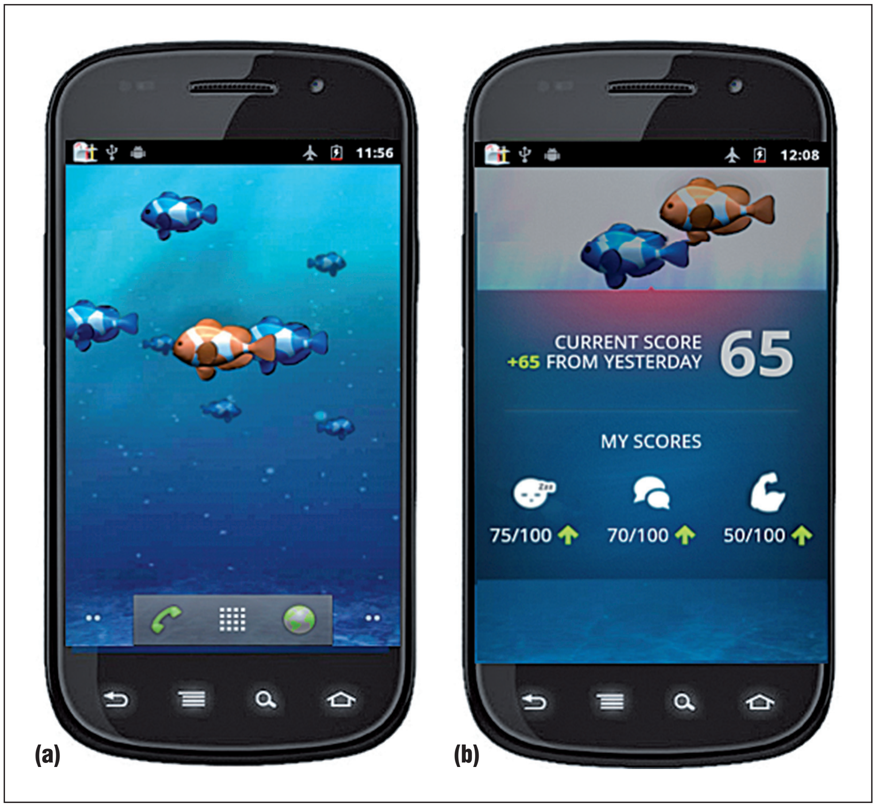


Figure 1. The BeWell mobile health app: (a) the ambient display on the smartphone's wallpaper screen and (b) the sleep, social, and activity scores.

recognition technology to answer questions, make recommendations, and issue requests to other services. As a result, smartphones represent the first truly mobile ubiquitous computing device.

We've developed several sensing apps that exemplify advances in mobile HCI and that aim to exploit big data to scale human-behavioral modeling. (A broader survey of smartphone sensing appears elsewhere.[1])

BeWell: A Mobile Health App

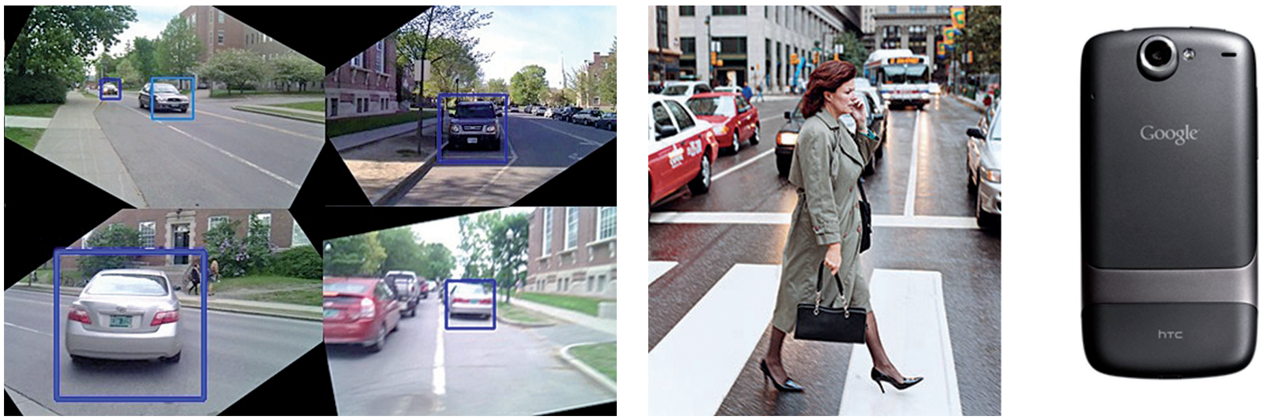
Smartphone sensing, combined with persuasive feedback techniques, is enabling a new generation of mobile health apps that can automatically monitor and promote multiple aspects of physical and emotional well-being. The BeWell app (for Androids) continuously tracks user behaviors along three distinct health dimensions without requiring any user input—the user simply downloads the app and uses the phone as usual (see https://play.google.com/store/apps/details?id=org.bewellapp).

Classification algorithms run directly on the phone to automatically infer the user's sleep duration, physical activity, and social interaction. In addition to classifying activities that influence health, BeWell also computes a weighted score between 0 and 100 for each dimension. A score of 100 indicates that the user is matching or exceeding recommended guidelines (averaging eight hours of sleep per day, for example).

The user's physical activity is classified as walking, stationary, or running, and inferences are used to estimate a daily Metabolic Equivalent of Task value. We rely on the Centers for Disease Control and Prevention's physical activity guidelines to parameterize the scoring system. Sleep monitoring estimates the user's sleep duration over a 24-hour period without the user having to do anything special with the phone; the user simply follows his or her normal behavior at bedtime—whether that's recharging the phone or leaving it on a table somewhere in the home. Our scoring system uses the guidelines for sleep duration provided by the National Sleep Foundation—too much or too little sleep is unhealthy and is thus reflected in the score.

We detect changes in social isolation based on the total duration of ambient speech during a day. This is estimated from the output of a speech/nonspeech classifier using the phone's microphone. (Audio isn't recorded on the phone or the cloud for privacy reasons.) We rely on studies that connect social isolation and social support to psychological well-being, with low levels being linked with symptoms such as depression. We experimentally develop a scoring system using field trials to determine the typical daily quantities of speech encountered by people within the study.[2] In addition to conversation, the social interaction dimension also considers the use of social applications on the phone (such as Facebook, voice calls, and email) when computing a composite sociability score for the user.

BeWell can run in a stand-alone mode on the phone or can interwork with the cloud to store longitudinal data patterns. BeWell promotes improved behavioral patterns via persuasive feedback as part of an animated aquatic ecosystem rendered as an ambient display on the smartphone's wallpaper screen (see Figure 1a). The speed of the large orange clown fish mirrors the user's activity, while the number of small blue fish reflect the user's level of social interaction with other people. Finally, the ocean's ambient lighting conditions indicate the user's sleep duration the previous night. Users can passively glance at the visualization of their health dimensions and reflect on how they're doing. At any time, the user can also view his or her current scores and the increase (up arrow) or decrease



(a) (b)

Figure 2. The WalkSafe app (a) offers real-time detection of the front and back views of cars, noting when a car is approaching or moving away from a user on the phone. (b) Each video frame is preprocessed to compensate for the phone tilt.

(down arrow) from the previous computed score (see Figure 1b).

The BeWell+ cloud service lets the user not only view his or her scores but also compare them with other BeWell users as a social network. The phone system sends targeted messages to users to encourage them to get back on track should the system note a low score.

Many of the challenges of building BeWell related to developing low-energy sensing capabilities, feature engineering, and the accurate classification of health dimensions without limiting the phone's battery lifetime or usability.

WalkSafe: Pedestrian Safety App

Research in social science has shown that mobile-phone conversations distract users, affecting pedestrian safety. For example, a mobile-phone user deep in conversation while crossing a street is generally more at risk than other pedestrians not engaged in such behavior.[3] We developed WalkSafe, an Android app for people who walk and talk.[4]

WalkSafe uses the smartphone's back camera to detect vehicles approaching the user, alerting the user of any potentially unsafe situations. More specifically, WalkSafe uses machine-learning algorithms implemented on the phone to detect moving vehicles. It also exploits phone APIs to save energy by running the vehicle-detection algorithm only during active calls.

The WalkSafe app offers real-time detection of the front and back views of cars, noting when a car is approaching or moving away from the user, respectively (see Figure 2). It alerts the user using sound and phone vibration.

The core WalkSafe car-detection technology is based on image-recognition algorithms. Image recognition is a computationally intensive process that, if not carefully designed, can easily drain the smartphone's computational resources and batteries. To address this, WalkSafe bases its vehicle recognition process on a model that's first trained offline and then uploaded to the phone and used for online vehicle recognition, which runs automatically whenever there's an ongoing phone call. WalkSafe activates the smartphone's camera and captures video of the surroundings. Each video frame is preprocessed to compensate for the phone tilt (as shown in Figure 2b) and illumination variations, and is then analyzed by the decision tree model built during the offline training phase. If the decision tree detects a car in the picture, it triggers an alert to warn the user of possible danger. WalkSafe is able to detect when a car is approaching at 30 miles per hour.

NeuralPhone: A Brain-to-Smartphone Interface

There's a growing interest in new hands-free interfaces for smartphones based on voice and face recognition systems. We developed the EyePhone, which lets the user select and activate applications with the blink of an eye.[1] We then wondered if a thought could also drive a smartphone application—and it turns out it can.

Until recently, devices for detecting neural signals were costly, bulky, and fragile. We developed the NeuralPhone, which uses neural signals to drive applications on the iPhone using inexpensive off-the-shelf wireless electroencephalography (EEG) headsets (see Figure 3).[1] We demonstrated a brain-controlled address book dialing app, which works like a P300 speller designed for brain-computer interfaces. The phone flashes a sequence of photos of contacts from the address book, and a P300 brain potential is elicited when the flashed photo matches the person whom the user wishes to dial. EEG signals from the headset are transmitted wirelessly to an iPhone, which natively runs a lightweight classifier to discriminate P300 signals from noise. When a person's contact photo triggers a P300, his or her phone number is automatically dialed. NeuralPhone breaks new

SMARTPHONES

Figure 3. The NeuralPhone app. It breaks new ground as a brain-to-smartphone interface for pervasive computing by letting users dial a smartphone contact via thoughts.

ground as a brain-to-smartphone interface for pervasive computing.

Community Similarity Networks: Big Sensor Data

Big data presents both challenges and opportunities. Today, we're seeing the emergence of new mobile health, well-being, and self-quantification apps that can automatically generate large numbers of sensor data streams. These streams are stored on phones and in the cloud for further mining, sharing, and visualization.

The BeWell application, for example, doesn't send raw data to back-end servers; rather, it uploads features, inferences, scores, and usability data to the cloud if the user opts to store and view longitudinal data. A typical BeWell user will upload 20 Mbytes of data per day when his or her phone is charging and connected to the Internet. Continuous sensing applications will gain popularity, producing terabytes of data that will need to be stored and processed in the cloud and potentially shared on social networks.

As the user population of smartphone sensing apps grows, the differences between people will quickly degrade the accuracy of the classification system—we call this the *population diversity problem.*[5,6] For example, how a young child walks differs greatly from how an elderly person walks, so the same model can't be used. To address this problem, we developed Community Similarity Networks (CSN), a classification system that can be incorporated into smartphone sensing apps to address the challenge of building robust classifiers for diverse populations.

The conventional approach to classification in mobile sensing is to use the same classification model for all users. Using CSN, we construct and continuously revise a personalized classification model for each user over time. Typically, personalized models require all users to provide hand-annotated examples of them performing certain activities while their devices gather sensor data (that is, label the data). This is both burdensome to the user and wasteful, because multiple users often collect nearly identical data, yet the training of each model occurs in isolation. CSN's key contribution is that it makes the personalization of classification models practical by significantly lowering the burden on the user by combining crowd-sourced data and leveraging networks that measure the similarity between users.

Under CSN training, classifiers become a networked process in which the effort of individual users benefits everyone. However, the use of crowd-sourced data must be done carefully. Crowd-sourced data must only selectively be used during training, so the resulting model is optimized for the person using the model. CSN solves this problem by maintaining similarity networks that measure the similarity between people within the broader user population. We do this by proposing three different similarity metrics (physical, lifestyle/behavioral, and purely sensor-data-driven metrics) that measure different aspects of interperson diversity that influence classifier performance. The CSN model training phase uses forms of boosting and co-training to let these different types of similarity each contribute to improving the personalized classifier's accuracy.

TOWARD COGNITIVE PHONES

By pushing intelligence to the phone in the form of classification models, we can infer human behavior and context. We can exploit big data to build more accurate and robust classification systems. Because people carry their phone as they navigate through the day, phones are well situated to go beyond simple inference of classes by building up knowledge of the user's life patterns and choices. What if a phone could not only build lifelogs but also predict outcomes and assist the user? We argue the next step in the evolution of the phone is the cognitive phone.

It's easy to imagine that the next generation of mobile health applications will not only track the user's physical, cognitive, and mental health but also use data analytics and prediction to model trends in the data. Thus, application-specific evidence—such as progressive social isolation, inactivity, and sporadic sleep patterns—could help predict the manic and depressive phases of someone suffering from a serious mental illness, such as a bipolar disorder. If the phone could accurately predict this change in health, could it also intervene to help the patient?

Another example relates to using sensor fusion and prediction on the phone. Changes in speech production are one of many physiological changes that happen during stressful situations. We recently developed the StressSense app on a quad-core Android phone,[7] which unobtrusively recognizes stressors from the human voice using the smartphone microphone. Microphones, embedded in mobile phones, provide the opportunity to continuously and noninvasively monitor stress levels in real-life situations.

Imagine a cognitive phone capable of fusing StressSense output (that is, a robust classification of stressors) with other phone data such that it could correlate and attribute stressors with people, meetings (from your phone's calendar), your health (correlations with BeWell), events (deadlines), and places (your manager's office). Consider, for example, that the phone's calendar overlays a simple color code representing your stress levels so you can visually understand at a glance what events, people, and places in the past—and thus likely in the future—aren't good for your mental health. Armed with this knowledge, the cognitive phone could help you avoid stressful situations by, for example, rearranging your calendar to avoid certain people, events, and locations. If your phone could understand your DNA, it might also offer suggestions to improve your overall well-being.

These motivational scenarios align with many of the open challenges in AI. An enduring difficulty AI researchers face is figuring out how to make systems more flexible, adaptable, and extensible. The development of cognitive phones will require tackling these challenges in the domain of human behavior as well as providing context recognition that works at the population level and throughout a user's lifetime. For example, in a mobile phone-based sensing app, the human user is always present and hence potentially able to provide helpful input, such as labels for data. However, an intelligent system will use this human resource sparingly and only when the potential information to be gained outweighs the inconvenience of interrupting the user.

Similarly, cognitive phones will seek to intelligently combine information from different sources, not by generic data pooling but by leveraging known relationships between human behavior at the group and individual levels. The phone would require a reasoning framework that considers multiple objectives and makes different types of decisions based on user needs such as whether to intervene (in the case of a patient relapse), offer a suggestion (perhaps reorganizing the user's calendar based on measured stressors), or taking action (such as ordering and paying for a latte in advance).

Each app discussed here pushes intelligence to the phone to infer different aspects of human behavior and context. The cellphone's rapid evolution into the smartphone has been breathtaking; the next evolutionary step should realize the cognitive phone.

ACKNOWLEDGMENTS

The following faculty, postdocs, and students (listed in alphabetic order) helped develop the smartphone sensing research presented here: Saeed Abdullah, Shahid Ali, Ethan Berke, Giuseppe Cardone, Gokul T Chittaranjan, Antonio Corradi, Afsaneh Doryab, Shane Eisenman, Kristof Fodor, Denise Frauendorfer, Daniel Gatica-Perez, Shaohan Hu, Nic Lane, Mu Lin, Hong Lu, Emiliano Miluzzo, Matthew Mukerjee, Mirco Musolesi, James Oakley, Wei Pan, Michela Papandrea, Mashfiqui Rabbi, Rajeev D.S. Raizada, Andy M. Sarroff, Lorenzo Torresani, Tianyu Wang, Ye Xu, Xiaochao Yang, and Chuang-wen You.

REFERENCES

1. N.D. Lane et al., "A Survey of Mobile Phone Sensing," *IEEE Comm. Magazine*, Sept. 2010, pp. 140–150.
2. M. Rabbi et al., "Passive and In-situ Assessment of Mental and Physical Well-Being Using Mobile Sensors," *Proc. 13th ACM Int'l Conf. Ubiquitous Computing* (Ubicomp 11), ACM, 2011, pp. 385–394.
3. J. Hedlund, "Pedestrian Traffic Fatalities by State: 2010 Preliminary Data," Governors Highway Safety Assoc., 2010, www.ghsa.org/html/publications/pdf/spotlights/spotlight_ped.pdf.
4. T. Wang et al., "WalkSafe: A Pedestrian Safety App for Mobile Phone Users Who Walk and Talk While Crossing Roads," *Proc. ACM HotMobile*, ACM, 2012; www.hotmobile.org/2012/papers/HotMobile12-final66.pdf.
5. N.D. Lane et al., "Enabling Large-Scale Human Activity Inference on Smartphones Using Community Similarity Networks (CSN)," *Proc. 13th Int'l Conf. Ubiquitous Computing* (Ubicomp 2011), ACM, 2011, pp. 355–364.
6. S. Abdullah, N. Lane, and T. Choudhury, "Towards Population Scale Activity Recognition: A Scalable Framework for Handling Data Diversity," to be published in *Proc. 26th Conf. Artificial Intelligence* (AAAI 12), 2012.
7. H. Lu et al., "StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones," to be published in *Proc. 14th Int'l Conf. Ubiquitous Computing* (Ubicomp 2012), ACM, 2012.

Andrew Campbell is a professor of computer science at Dartmouth College. Contact him at campbell@cs.dartmouth.edu.

Tanzeem Choudhury is an associate professor of information science at Cornell University. Contact her at tanzeem.choudhury@cornell.edu.