Affect Sensing on Smartphone - Possibilities of Understanding Cognitive Decline in Ageing Population.

Rajib Rana, John Reilly, Raja Jurdak *Senior Member, IEEE*, Wen Hu *Senior Member, IEEE*, Xue Li, *Member, IEEE*, Jeffrey Soar

***Abstract*—Due to increasing sensing capacity, smartphones offer unprecedented opportunity to monitor human health. Affect sensing is one such essential monitoring that can be achieved on smartphones. Information about affect can be useful for many modern applications. In particular, it can be potentially used for understanding cognitive decline in aging population. In this paper we present an overview of the existing literature that offer affect sensing on smartphone platform. Most importantly, we present the challenges that need to be addressed to make affect sensing on smartphone a reality.**

I. Introduction

Smartphones offer unprecedented opportunity to monitor personal health. These phones are built with plethora of sensors, which offer multimodal sensing of physiological, psychological or emotional and behavioural functions of human being. Commonly used sensors on smartphone include three axis accelerometer, gyroscope, microphone, camera, GPS receiver, proximity sensor, luxmeter, temperature and humidity sensor. Besides these physical sensors, information driven from various services available on smartphone are also used for profiling human functions. The commonly used services include, call and SMS logging and Bluetooth and Wi-Fi scanning. Last but not the least, very recently, interaction patterns, such as typing speed on the phone keyboard, swiping speed on the touch screen, various "key" pressing frequencies (such as, "Backspace Key", "Enter Key", "Special Symbol" etc.), maximum text length, erased text length and touch count are also used for profiling various human functions.

Smartphone sensing is particularly suitable for aging population as it does not introduce an additional sensing device. This group do not prefer to carry additional sensing device from the sense of being monitored. However, they need to carry the phone for communicating with peer, family members and for making many other important communications. Therefore, smartphones can potentially collect rich information about their various functions. In this paper we focus on the feasibility of smartphone sensing for understanding affect. We also discuss the feasibility of inferring cognitive status of the aging population from the available information about affect. We organize the rest of the paper as follows. In the next section (Section II) we define affect, discuss commonly used affect models and its application in understanding cognition in aging population. In Section III we discuss various modalities for affect sensing and present positive and negative aspects and challenges in each modality. In Section IV, we discuss and compare studies pertaining to various modalities discussed in Section III. Finally, in Section V we conclude and discuss potential future directions.

Rajib Rana, Raja Jurdak and Wen Hu are with the department of Computational Informatics, CSIRO, Australia (firstname.lastname@csiro.au)

John Reilly is with the Mental Health Service Group in the Townsville Hospital, Australia (email: John.Reilly@health.qld.gov.au)

Xue Li is with the School of Information Technology and Electrical Engineering, University of Queensland (e-mail: x.li@uq.edu.au).

Jeffrey Soar is with the School of Management and Enterprise, University of Southern Queensland (e-mail: Jerey.Soar@usq.edu.au).

II. Affect Sensing and Its Application in Understaning Cognition in Aging Population

A. Affect and Affect Models

A person's affect is the expression of emotion or feelings displayed to others through facial expressions, hand gestures, voice tone, and other emotional signs such as laughter or tears.

In psychology there a number of extensively validated models are proposed to describe affect. Within the scope of the paper we have considered four different models, which are Circumplex mood model [1], Ekman's six basic categories [2], Positive and Negative Affect Schedule (PANAS) [3], [4] and finally, the Big-Five Model [5].

The Circumplex mood model describes affect in only two dimensions: the pleasure dimension and the activeness dimension. The pleasure dimension measures how positive or negative one feels and the activeness dimension measures whether one is likely to take an action under the mood state. It has been proven that users are consistently able to place discrete affect in these two dimensions.

Ekman's six basic categories: happiness, sadness, fear, anger, disgust, and surprise are very popular in describing affect. This approach is intuitive and matches peoples' daily experience well.

The third model PANAS assumes that it is possible to feel good and bad simultaneously. Therefore, it tracks the positive and negative affects separately. However, the complexity of PANAS model makes its integration difficult in real life automated applications.

Finally, the Big-Five Model uses a hierarchical model to describe personality traits in five dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. This model is comprehensive and applicable across various cultures.

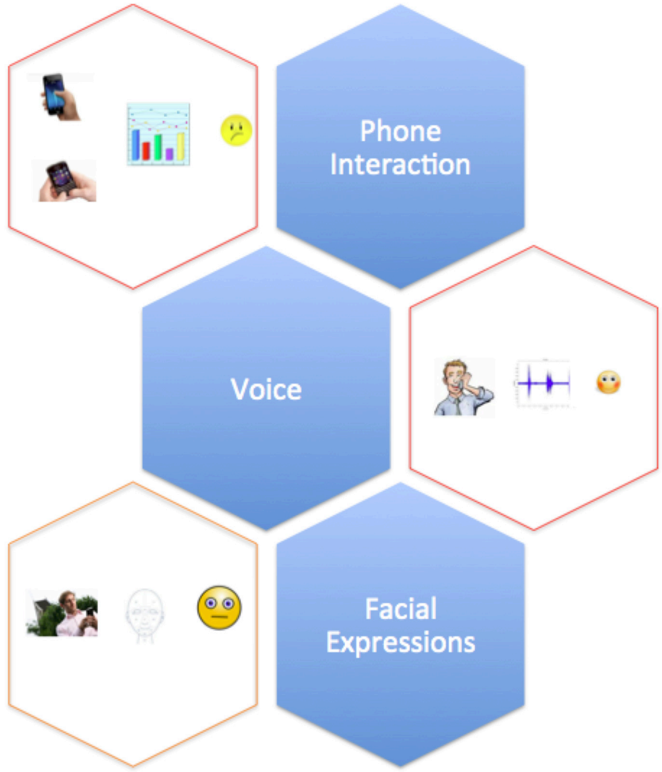

Fig. 1: Affect Sensing on Smartphone.

B. Affect Sensing for Coginitive Status

Cognitive decline is a strong predictor for many neurodegenerative disorders in aging population. Studies([6], [7]) have shown that human cognition and affect has many dependencies. Therefore, affect sensing on the phone can facilitate the understating of the interplay between affect and cognition. Thus, it can potentially help to predict the cognitive decline by identifying specific patterns in the affect.

III. Affect Sensing on Smartphone

The existing methods of affect sensing on smartphone can be grouped into three categories: 1. Affect sensing using acoustic data, 2. Affect sensing from phone interactions/usage, and 3. Affect sensing from facial expressions. In Figure 1 we have illustrated the working principle of these three categories of affect sensing. Below we describe these categories in details.

A. Affect Sensing from Acoustic Data

Large body of work can be found in the literature where researchers have used voice for affect sensing [8], [9]. However, many of these algorithms require ample processing capacity, which cannot be afforded on the smartphone platforms. Only a small number of studies can be found where affect sensing has been done on the smartphone using acoustic samples. Amongst these studies [10] is significantly advanced. In this work researchers have determined stress in both indoor and outdoor environments. However, it is not clear from the paper that how data was collected from outdoor. It would be much challenging if the stress is determined when the user is having conversation over phone - especially in presence of background noise. The other challenge would be getting access to data. It would be complicated to get ethical clearance to access conversation on the phone. Last but not the least, the publicly available datasets only use sentences pronounce by actors. However, most of these recordings are not on phone, rather on high-end microphone, so the sound does not have similar attributes if it is recorded using the phone microphone. Due to automatic gain control and noise cancellation recording using the phone microphone could potentially loose important acoustic attributes.

B. Affect Sensing using Phone Interactions/Usage

A number of studies have shown in the past that typing speed on computer keyboard and mouse click speed have correlation with human behaviour [11]–[17]. As a follow-up of those outcomes, researchers seek to find correlation of smartphone usages with affect. The usage under consideration include, number of SMS sent, number of calls received and made, duration of calls etc. The interaction is composed of number of presses on backspace, number of presses on Enter key etc. In addition, social interaction patterns mined from Bluetooth and Wi-Fi scan and GPS traces are used to model affect.

Affect sensing using smartphone interactions/usage have higher degree of non-invasiveness compared to audio and video based assessment, however, they have limited correlation with affect. Therefore, accuracy is generally low (see Table I).

C. Affect Sensing using Facial Expressions

A large number of studies in the past have shown that facial expression is strongly correlated with mental wellbeing [18]–[20]. For example, it has been reported that in the depressed episodes patients pose neutral expression compared to any other expressions. However, extracting facial expression on the smartphone is a power intensive operation. It requires face area detection and then expression classification. In addition, the existing algorithms work well in controlled setting, without the change in lighting condition, occlusion, however, in natural setting it is a challenging task.

Performance of affect sensing using this method can be significantly improved if spontaneous facial expressions can be captured though out the day. A related application called Face log [21] offers the opportunistic face image capturing through the day via various mobile devices, including Google glass, smart phone etc. Such an application with expression extraction facilities would be highly useful for inferring affect. This spontaneous captures will contain more latent information instead of few minutes of self-initiated recording, where most of the people behave differently than usual.

TABLE I: Comparison of Existing Studies.

Cat.	Emotions Analyzed	Ground Truth	Sample Size	Sources of features	Results	Method	Future Work
Phone Interactions/Usage	Happiness, Surprise, Anger, Disgust, Sadness, Fear, Neutral.	Tweet.	One participant over two weeks.	Keystrokes, Touchscreen parameters, Discomfort index, Location, Time, Weather	67.52% on average, the best accuracy for happiness, surprise and neutral state [22].	Weka s/w suite.	Investigate speed or intensity of a continual touch (e.g., drag operation) as a feature.
	Pleasure and activeness (Circumplex mood model).	Mood Journal.	32 participants over two months.	Application usage, Phone calls, email messages, SMS, Web browsing history, and Location change.	Initial accuracy 66% which gradually grows to 93% over two months of training [23].	Multi-Linear Regression	Future study includes how mood models can be trained using data from multiple people while still providing sufficient guarantees of privacy to each user.
	Happiness.	Self-reported surveys about personality traits (Big Five) and daily happiness.	117 participants over eight weeks.	Call logs, SMS logs, Proximity data, obtained by scanning near-by phones and other Bluetooth devices every five minutes.	80.81% [24].	Random Forest classifier.	Investigation of predictive capacity of the SMS data.
	Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness to Experience.	TIPI questionnaire.	83 participants over eight months.	App logs, Call logs, SMS logs, Bluetooth Logs.	75.9% [25].	Decision Tree and SVM Classiifer with RBF Kernel.	Analyze other modalities, such as accelerometer and GPS logs.

Cat.	Emotions Analyzed	Ground Truth	Sample Size	Sources of features	Results	Method	Future Work
Acoustic Data	Anger, Boredom, Disgust, Fear, Happiness, Sadness, Neutral.	Metadata from Database.	Not mentioned	Speech samples from four datasets: Belfast, Reading-leeds, CREST-ESP, Berlin.	81.5% [26].	Features: Formants, MFCC, Centre of Gravity, Sprectum Central Moments, Standard Deviation, Skewness, Kurtosis, Glottal Pulses. PCA is used for feature dimension reduction Classfer: custom defined.	Computational load analysis.
	Stress.	Galvnic Skin Reistance (GSR) sensor.	14 participants, 10 females and 4 male.	Speech samples from four datasets: Belfast, Reading-leeds, CREST-ESP, Berlin.	81% and 76% accuracy for indoor and outdoor, respectively [27].	Features: Pitch (F0); Classifier: gaussian Mixture Model (GMM).	Using self-train model for speaker adaption and the supervised adaption model for environment adaption.
	Anger, Happiness, Neutral, Sadness.	Metadata from database.	Five males and five females were invited to simulate seven kinds of emotions with 10 different sentences. There were 493 speech utterances, in which 286 speech samples were of female voices and 207 were of male voices.	No data was collected using smartphone. Berlin dataset was used.	86% [28].	Features: Pitch (F0), Pulses, Voice, Jitter, Shimmer, Harmonics; Classifier: neuro-fuzzy network with a weighted fuzzy membership function.	Not mentioned specifically. However, authors suggest its use as an assistant tool for professional counseling, for dynamically monitoring clients emotional changes and then providing them with the appropriate online help.

Cat.	**Emotions Analyzed**	**Ground Truth**	**Sample Size**	**Sources of features**	**Results**	**Method**	**Future Work**
Acoustic Data	Stress.	PANAS questionnaire.	35 participants, over 4 months.	Acceleration and GPS traces, Call logs, Battery levels, Address book, Microphone.	61% [10].	Smartphone features: audio, physical activity, social interaction. Heart Rate Variability features: time domain features, nonlinear features, frequency domain features. Classifier: multinomial logistic regression.	Use advanced psychological sensors such as Empactica, Affectiva Q sensor etc.
	Stress.	Exam period was used as an indicator for stress and otherwise.	7 participants, over four weeks.	GPS, WiFi traces, Call logs, Bluetooth, SMS.	Overall behavior modification was used in lieu of accuracy [29].	Call behavior, SMS behavior, social interaction behavior and places of interest visiting behavior were used as features.	Exact interpretation and generalization requires further work and larger scale experiments.
Facial Expression	Facial Expressions.	Not mentioned.	Not mentioned.	Facial image, video.	Not mentioned [30].	Open CV library.	Not mentioned.

Facial expressions will also be very useful for affect sensing if expressions can be correlated with context. Studies in the past has verified the success of this method. Patients were showed movie clips with different emotional contents and corresponding responses in facial expression were recorded. The initial hypothesis was that the patients' expressions would be mostly sad, however, the results showed that the patients mostly posed neutral facial expression. This study was conducted in the laboratory using computer. Facilitating such an assessment on the smartphone would be heavily beneficial, since patient do not need to travel to the laboratory. Furthermore, this method can be even advanced, if instead of showing predefined content, if the content that the patient is watching can be automatically categorized.

IV. Comparison of various Affect Sensing Methods

In this section we compare the performance of the three categories of the affect sensing methods discussed in Section III. As a basis of comparison we nominate a number of aspects, such as emotion analysed, ground truth, sample size, source of features, accuracy, method and future work.

Emotion Analyzed: This will provide the readers the information on what type of emotions are considered. Some researchers may be interested in the performance of inferring particular emotion, such as, happiness whereas others have considered broader range of emotions (e.g., disgust, boredom and others). This listing will help reader to navigate to a particular study that considers the emotion of interest. In addition, this will also form a basis of comparison amongst different studies. Generally, the method scoping higher number of emotions is considered comprehensive.

Ground Truth: This will inform the reader the various ways of validating sensor driven inference of affect.

Sample Size: This includes number of participants and duration of experiments. This is very important to evaluate the significance of the results achieved by a particular study.

Source of Features: This is an interesting aspect of comparison. This reports the diverse range of features that can be used for emotion inference.

Results: This is an obvious means of comparison. However, a study needs to evaluated jointly by the results and the sample size.

Method: Under this heading features and classification methods used within the corresponding study will be discussed. This is important to potentially compare various features and classification methods.

Future Direction: Under this heading open questions or future work directions discussed in various studied are provided. Researchers interested in extending any of the existing works can use these as a starting point of development. In Table I we briefly summarize the comparison of various studies under the three categories considered in the paper.

V. Conclusion and Future Challenges

In this paper we have discussed various affect sensing methods that have been developed on the smartphone platform. We have discussed the potential of affect sensing for predicting cognitive decline in aging population. We also presented a brief comparison amongst these methods which will assist the readers quickly grasp the research landscape. In summary, amongst these methods, the technique of extracting latent information about affect from smartphone usage is quite recent and promising, however, the accuracy needs further improvement. In addition, since this is a fairly new approach, long-term clinical studies need to be carried out to demonstrate the correlation of various interactions with affect. The limitations of using voice and facial expressions for affect sensing are two fold: first, it requires substantial amount of system resource which may compromise battery life. Second, opportunistic sampling is sought instead of predefined sampling. For example, facial expressions need to be captured opportunistically when people are engaged in daily routine. Similarly, voice needs to be assessed when people are engaged in phone conversation. Opportunistic sampling of voice and facial expressions will however pose challenges of extracting features from voice with background noise and evaluating facial expression under changing lighting condition, occlusion, pose variation etc, respectively. The theory of Compressive Sensing(CS) have recently attract research interest as it offers accurate classification [31]–[35] and reconstruction [36]–[41] within resource constraints. The implement ion of CS theory for classification is commonly known as Sparse Random Classifier (SRC). The most desirable attribute of SRC is that it does not require feature extraction and training, which is a critical requirement of most of the classical classification algorithms. SRC has been used in facial expression classification, abnormal event classification, gait classification, physical activity classification and so on. It will be interesting to find out the accuracy versus resource consumption capacity of SRC on smartphone platform. Last but not the least fusion of various categories will be another interesting area to explore to achieve better accuracy.

References

[1] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[2] P. Ekman, "Universals and cultural differences in facial expressions of emotion." in *Nebraska symposium on motivation*. University of Nebraska Press, 1971.

[3] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales." *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[4] J. R. Crawford and J. D. Henry, "The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample," *British Journal of Clinical Psychology*, vol. 43, no. 3, pp. 245–265, 2004.

[5] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, 1992.

[6] S. Schachter and J. Singer, "Cognitive, social, and physiological determinants of emotional state." *Psychological review*, vol. 69, no. 5, p. 379, 1962.

[7] P. N. N. Inferences, "Feeling and thinking," *AMERICAN PSYCHOLOGIST*, p. 1S1, 1980.

[8] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 2, pp. 293–303, 2005.

[9] B. Schuller, R. J. Villar, G. Rigoll, M. Lang *et al.*, "Meta-classifiers in acoustic and linguistic feature fusion-based affect recognition," in *Proc. ICASSP*, vol. 5, 2005, pp. 325–328.

[10] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.

[11] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation computer systems*, vol. 16, no. 4, pp. 351–359, 2000.

[12] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," *ACM Transactions on Information and System Security (TISSEC)*, vol. 5, no. 4, pp. 367–397, 2002.

[13] D. Gunetti and C. Picardi, "Keystroke analysis of free text," *ACM Transactions on Information and System Security (TISSEC)*, vol. 8, no. 3, pp. 312–347, 2005.

[14] P. S. Dowland and S. M. Furnell, "A long-term trial of keystroke profiling using digraph, trigraph and keyword latencies," in *Security and protection in information processing systems*. Springer, 2004, pp. 275–289.

[15] R. V. Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *International Journal of Biometrics*, vol. 1, no. 1, pp. 81–113, 2008.

[16] M. Pusara and C. E. Brodley, "User re-authentication via mouse movements," in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*. ACM, 2004, pp. 1–8.

[17] A. A. E. Ahmed and I. Traore, "Detecting computer intrusions using behavioral biometrics." in *PST*, 2005.

[18] J. F. Cohn, "Foundations of human computing: facial expression and emotion," in *Proceedings of the 8th international conference on Multimodal interfaces*. ACM, 2006, pp. 233–238.

[19] M. Pantic and M. S. Bartlett, "Machine analysis of facial expressions," 2007.

[20] P. Ekman and W. V. Friesen, "Facial action coding system," 1977.

[21] C. Yoo, J. Shin, I. Hwang, and J. Song, "Facelog: capturing user's everyday face using mobile devices," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 2013, pp. 163–166.

[22] H. Lee, Y. S. Choi, S. Lee, and I. Park, "Towards unobtrusive emotion recognition for affective social communication," in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*. IEEE, 2012, pp. 260–264.

[23] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 389–402.

[24] A. Bogomolov, B. Lepri, and F. Pianesi, "Happiness recognition from mobile phone data," in *Social Computing (SocialCom), 2013 International Conference on*. IEEE, 2013, pp. 790–795.

[25] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's who with big-five: Analyzing and classifying personality traits with smartphones," in *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*. IEEE, 2011, pp. 29–36.

[26] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese, and A. Sciarrone, "Gender-driven emotion recognition through speech signals for ambient intelligence applications," *Emerging Topics in Computing, IEEE Transactions on*, vol. 1, no. 2, pp. 244–257, Dec 2013.

[27] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, "Stresssense: Detecting stress in unconstrained acoustic environments using smartphones," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: ACM, 2012, pp. 351–360. [Online]. Available: http://doi.acm.org/10.1145/2370216.2370270

[28] M. Viswanathan, Z.-X. Zhang, X.-W. Tian, and J. S. Lim, "Emotional-speech recognition using the neuro-fuzzy network," in *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC '12. New York, NY, USA: ACM, 2012, pp. 96:1–96:5. [Online]. Available: http://doi.acm.org/10.1145/2184751.2184863

[29] G. Bauer and P. Lukowicz, "Can smartphones detect stress-related changes in the behaviour of individuals?" in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, March 2012, pp. 423–426.

[30] V. Gay and P. Leijdekkers, "Design of emotion-aware mobile apps for autistic children," *Health and Technology*, pp. 1–6, 2013. [Online]. Available: http://dx.doi.org/10.1007/s12553-013-0066-3

[31] B. Wei, M. Yang, R. K. Rana, C. T. Chou, and W. Hu, "Distributed sparse approximation for frog sound classification," in *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM, 2012, pp. 105–106.

[32] J. Xu, S. Denman, S. Sridharan, C. Fookes, and R. Rana, "Dynamic texture reconstruction from sparse codes for unusual event detection in crowded scenes," in *Proceedings of the 2011 joint ACM workshop on Modeling and representing events*. ACM, 2011, pp. 25–30.

[33] S. Sivapalan, R. K. Rana, D. Chen, S. Sridharan, S. Denmon, and C. Fookes, "Compressive sensing for gait recognition," in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*. IEEE, 2011, pp. 567–571.

[34] W. Xu, M. Zhang, A. A. Sawchuk, and M. Sarrafzadeh, "Co-recognition of human activity and sensor location via compressed sensing in wearable body sensor networks," in *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*. IEEE, 2012, pp. 124–129.

[35] S. W. Chew, R. Rana, P. Lucey, S. Lucey, and S. Sridharan, "Sparse temporal representations for facial expression recognition," in *Advances in Image and Video Technology*. Springer, 2012, pp. 311–322.

[36] R. Rana, W. Hu, T. Wark, and C. T. Chou, "An adaptive algorithm for compressive approximation of trajectory (aacat) for delay tolerant networks," in *Wireless Sensor Networks*. Springer, 2011, pp. 33–48.

[37] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*. ACM, 2010, pp. 105–116.

[38] R. Rana, W. Hu, and C. T. Chou, "Energy-aware sparse approximation technique (east) for rechargeable wireless sensor networks," in *Wireless Sensor Networks*. Springer, 2010, pp. 306–321.

[39] C. T. Chou, R. Rana, and W. Hu, "Energy efficient information collection in wireless sensor networks using adaptive compressive sensing," in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*. IEEE, 2009, pp. 443–450.

[40] Y. Shen, W. Hu, R. Rana, and C. T. Chou, "Non-uniform compressive sensing in wireless sensor networks: Feasibility and application," in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2011 Seventh International Conference on*. IEEE, 2011, pp. 271–276.

[41] Y. Shen, W. Hu, R. Rana, and C. T. Chou, "Nonuniform compressive sensing for heterogeneous wireless sensor networks," *IEEE Sensors*, vol. 13, no. 6, pp. 2120–2128, 2013.