

Demo: Situation-Aware Conversational Agent with Kinetic Earables

Shin Katayama[†], Akhil Mathur^{*}, Tadashi Okoshi[†], Jin Nakazawa[†], Fahim Kawsar^{*}

^{*}Nokia Bell Labs, Cambridge, UK, [†]Keio University, Kanagawa, Japan

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile computing systems and tools;

1 ABSTRACT

Conversational agents are increasingly becoming digital partners of our everyday computing experiences offering a variety of purposeful information and utility services. Although rich on competency, these agents are entirely oblivious to their users' situational and emotional context today and incapable of adjusting their interaction style and tone contextually. To this end, we present a first-of-its-kind situation-aware conversational agent on kinetic earable that dynamically adjusts its conversation style, tone, volume in response to users emotional, environmental, social and activity context gathered through speech prosody, ambient sound and motion signatures. In particular the system is composed of the following components:

- **Perception Builder:** This component is responsible for building an approximate view of user's momentary experience by sensing his/her 1) physical activity, 2) emotional state, 3) social context and 4) environmental context using different purpose-built acoustic and motion sensory models [4, 5].
- **Conversation Builder:** This component enables a user to interact with the agent using a predefined dialogue base, and for this demo, we have used Dialogflow [1] populated with a set of situation-specific dialogues.
- **Affect Adapter:** This component is responsible for guiding the adaptation strategy for the agent's response corresponding to the user's context, taking into account the output of the perception builder and a data-driven rule engine. We have devised a set of adaptation rules using multiple quantitative and qualitative studies that describe the prosody, volume and speed to shape agents response.
- **Text-to-Speech Builder:** This component is responsible for synthesising the agent's response in a voice that accurately reflects a user's situation using IBM Bluemix Voice service [2]. This synthesis process interplays various voice attributes, e.g., pitch, rate, breathiness, glottal tension etc. to transform agents voice according to the rule of the Affect Adapter.

The manifestation of these components is realised with a kinetic earable (the source of sensory signals and playback) and a Raspberry Pi Zero (execution platform). The kinetic earable, eSense, is an in-ear high definition wireless stereo wearable (Figure 1) instrumented with a microphone, a 6-axis inertial measurement unit and dual model

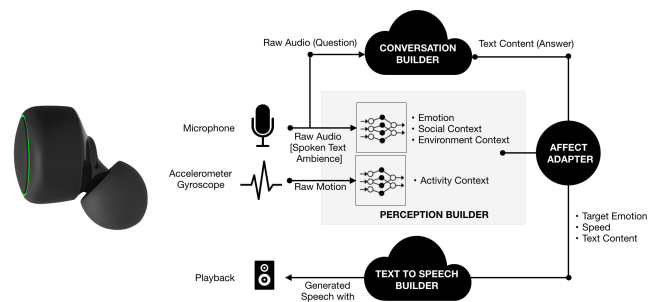


Figure 1: The end-to-end architecture of the situation-aware conversational agent with a kinetic earable that adjusts its interaction style contextually.

Bluetooth and Bluetooth Low Energy (BLE)[3]. These embodiments collectively enable eSense to offer three sensing modalities - audio, motion, and proximity that we have used for our sensory inferences.

Early experimental results demonstrate that our situation adaptation mechanism invariably elicits better and affective user experience in comparison to baseline conditions in different real-world settings.

2 DEMONSTRATION

The demonstration will showcase the conversational agent modelling a set of predefined lifestyle situations in the home and work context. The system will include an eSense kinetic earable that users will wear to converse with the agent, and a Raspberry Pi Zero that will host the computations, i.e., ingesting sensory signals from the earable, running the model to understand the situation, interacting with conversation builder to acquire the answers and finally converting text to speech for playback with affect adaptation.

ACKNOWLEDGEMENT

This work was supported (in part) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

- [1] 2019. Dialogflow. <https://dialogflow.com>. (2019). Accessed: 2019-04-17.
- [2] 2019. IBM Bluemix. <https://console.bluemix.net>. (2019). Accessed: 2019-04-17.
- [3] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.
- [4] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 213–224. <https://doi.org/10.1145/3242587.3242609>
- [5] Liangying Peng, Ling Chen, Zhenan Ye, and Yi Zhang. 2018. AROMA: A Deep Multi-Task Learning Based Simple and Complex Human Activity Recognition Method Using Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 74.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobiSys '19, June 17–21, 2019, Seoul, Republic of Korea

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6661-8/19/06.

<https://doi.org/10.1145/3307334.3328569>