She just doesn't understand me! Curing Alexa of her Alexithymia

Eoghan Furey[1] Juanita Blue [2]

[1] Letterkenny Institute of Technology, Letterkenny, Donegal, Ireland
[2] Ulster University, Magee campus, Derry, Northern Ireland
eoghan.furey@lyit.ie

Abstract. Voice Command Devices (VCD) such as the ubiquitous Amazon Echo have entered the lives and homes of people the world over. Although emotional recognition and digital empathy are idealistic goals within the development of AI and intelligent agents, the current technology available lacks outward emotional understanding and the personas contribute only Alexithymic (no understanding of emotion) responses. Despite extensive research by large multinational technological organizations, authentic human-like empathic interactions with intelligent agents have not yet been achieved. Consequently, users are lulled into a false sense of security where they believe that their emotions remain private.

This paper determines that despite Alexa's demonstrated lack of emotion and emotional understanding, Voice Command Devices such as the Amazon Echo have the ability to deduce emotions such as sadness through inferential data. This is displayed through responses to questions that offer the same information as those posed by health practitioners to establish potential cases of depression. This type of data paves the way for parent companies to effectively target future advertising and build EMOTOgraphic models. As users are presented with no indication of this by intelligent agents, most would be unaware that combined inferential data could be so revealing and potentially extremely profitable from a sales and marketing perspective. This potentially leads to great ethical and privacy concerns as intelligent agents such as Alexa are gradually and incrementally cured of Alexithymia indicators.

Keywords: Emotion, Amazon Echo, Alexa, Inferential data.

1 Introduction

Emotion recognition and digital empathy are visionary goals of research conducted within the areas of Human Computer Interaction (HCI) and Affective Computing [1]. Despite a myriad of data produced by sensors, detectors and Internet of Things (IoT) devices, and also the analysis conducted thereon, effective emulation of human emotion recognition remains a challenge for interactive artificially intelligent agents [2]. Typically, humans are emotive beings. Mood, feelings and emotions are woven into our interactions. They are conveyed by voice tone and intonation, facial expression, body

language and a plethora of other social behaviours. Despite all these cues, some individuals still struggle with understanding the emotion of others. Alexithymia is defined as "a personality construct characterized by the sub-clinical inability to identify and describe emotions in the self and others" [3]. Sufferers of Alexithymia display marked dysfunction in interpersonal relating, social attachment and emotional awareness [4]. Due to challenges in distinguishing and understanding emotions of others, the behaviours of alexithymics can appear un-empathic. The latest generation of intelligent agents housed by voice command devices (VCDs) may be able to support daily routines, but lack key indicators of emotional understanding and empathy [5].

VCDs offer a new paradigm for mainstream human interaction with artificial intelligence. These devices are becoming increasingly commonplace in domestic environments [6]. They provide access to intelligent personal assistant services that are designed to synthesize and streamline various aspects of daily life [7]. These devices are changing the way that people think about and interact with their environments. The influx of VCDs has the capacity to contribute a multitude of worthwhile and novel features that aid and improve the user experience when executing conventional tasks. The Amazon Echo [8] is currently one of the most popular and pervasive of these devices. The associated intelligent agent, 'Alexa' is capable of demonstrating human-like conversational ability, however her unemotive responses would allow all but the most primitive human minds to quickly establish that she is not human [8]. Alexa clearly neither recognises nor truly displays any emotion, therefore it could be posited that Alexa may suffer from Alexithymia! Whether the similarity in name is by design or coincidence, it is evident that a cure for this condition is a goal belonging to Amazon's developers. This research explores the means by which VCDs may infer an individual's emotional state, thereby reducing prevalence of Alexithymia in AI.

2 Background

This section provides background for the research, offering information that relates to the disciplines, technologies, models, associations and considerations concerned.

2.1 Affective Computing

Affective computing is an emerging interdisciplinary research area that combines various fields, ranging from artificial intelligence and natural language processing, to cognitive and social sciences [9]. A main goal of affective computing is to develop systems capable of adapting to users' emotions in order to produce more natural and efficient interaction. Thus, a central component of the field is emotion recognition based on a variety of measurements including facial expressions, speech, gait patterns and other metrics that are analysed using advanced pattern recognition techniques [10]. Although there have been great advancements in the field, only a few robust implementations have been presented or validated, thus adoption of affect-aware technology has been marginal [1].

2.2 Voice Command Devices

Becoming increasingly popular, these devices are typically interactive and function to gather metrics and support many aspects our lives [7]. While keyboard and pointer input have traditionally been the means of human-computer interaction, since 2012 there has been a significant drive [11] to enable hands free control, primarily by voice. Figure 1 depicts the operation of Voice Command Devices.

Fig. 1. VCDs connect to the Internet & Cloud

The Amazon Echo provides an artificially intelligent personal assistant referred to as 'Alexa', who possesses a range of functionality and 'skills' which enable voice activation and communication with numerous IoT devices. These include wearable health devices, shopping lists, streamed music services and calendars, amongst others [6]. Following use of the 'wake word' that activates the VCD, interactions and data harvested is subsequently stored in the parent company's cloud.

2.3 Emotion Recognition

Emotions are physiological, behavioural, and/or communicative reactions to stimuli that are cognitively processed and experienced [12]. They are often internally experienced through physiological changes such as increased heart rate or a tense stomach [13]. These physiological reactions may not be detected by others and are considered intrapersonal unless there is a verbal or non-verbal cue that indicates the internal state and these cues may be voluntary or involuntary [14]. When communicating, cues in verbal intonation and body language provide information to others relating to how they should react. For example, when someone exhibits behaviours associated with sadness, it is an indication that support is needed [12]. Humans typically learn through socialisation how to read and display emotions.

Emotion recognition describes the process of identifying emotion in humans. The cognitive mechanisms invoked in recognizing human emotion have been studied for hundreds of years. The fields of psychology and more recently those such as cognitive science, have developed a range of models in this area. The most recent models rely

on AI through use of signal processing, machine learning, natural language processing (NLP) and other modern implementations [15].

Within emotional recognition, signals detected by all five senses (sight, sound, touch, scent and taste) may be recruited to enable one human to recognize, process and understand the emotions of another human. However, typically the signals that are relied upon most heavily are those that are visual and produce sound [14]. Sight allows a human to process and understand facial expressions and body language relating to particular emotions. Hearing enables the emotion to be revealed through analysis of the language, pitch, tone and volume of a human voice [14]. These signals are transmitted to the brain, combined and analysed so that the emotion may be deduced. Humans have come to rely heavily on use of language to recognize the emotion of another since the advent of voice telecommunications and written correspondence [16]. Figure 1 displays a subset of the means by which a person may sense the emotion of another person.

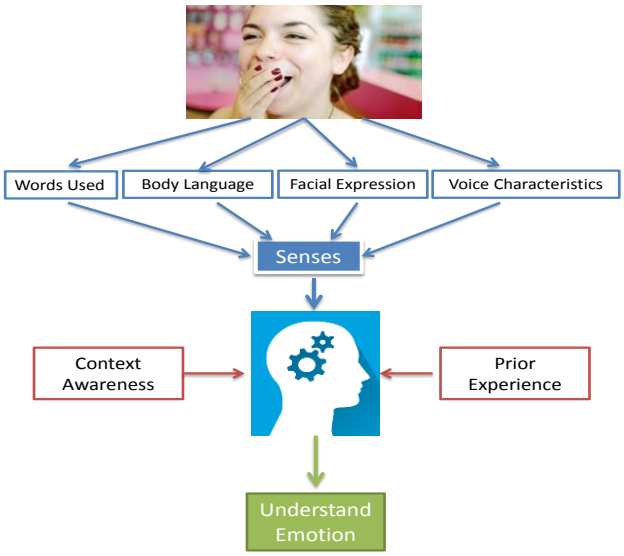

Fig. 2. Process of human emotion understanding

Currently, there is significant research in progress that focuses primarily on facial expressions, NLP and voice characteristics in a bid to help machines to learn to achieve emotion detection and recognition in a way that emulates human ability [17]. Research has also indicated that analysis of gait and posture may also reveal indicators of emotion [18]. Additionally, studies of human behaviour and social interactions may also contain clues relating to the emotional state of a person. EEG, ECG and MRI medical technologies facilitate the scanning of brains and bodies, these enable researchers and medical professionals to detect emotion associated neural signals that allow emotion detection and classification beyond the instinctive and learned mechanisms invoked by humans

[18]. Although there are many indicators that may be examined when using modern technology to detect and classify emotion, current successful models rely heavily on facial expression and textual analysis. The Amazon Echo device currently lacks the capacity to analyse facial expression and body language. Its resident intelligent agent, Alexa, must rely primarily on sound and interactive behavior [19].

2.3 Emotion & Voice Command Devices

Replicating the emotion detection capability of people within intelligent agents such as Alexa is an increasingly popular and innovative field of research [20]. Work in this area spans multiple related disciplines including AI, ML, NLP, Human Computer Interaction (HCI), Robotics and Affective Computing. Artificial sensors such as vision systems, microphones, thermo-detection devices, body electro sensors, magnetic resonance detectors, heart-rate monitors and sweat detection devices all produce data that may be analysed to elicit patterns that correspond to emotions [20]. Challenges arise when the number and variance of sensors or modalities is limited. However, inference may be used to 'fill in the gaps' and deduce emotions based on the data that is available. Popular systems that are invoking inferential data to overcome these challenges are embodied in Voice Command Devices such as the Amazon Echo [21]. The range of sources that contain data with which emotion may be detected by the Echo are displayed in Figure 3.

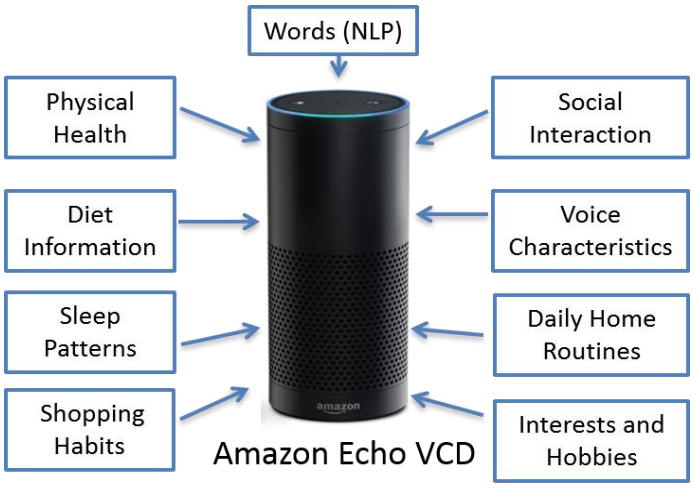

Fig. 3. Sources available to VCDs that enable them to infer emotion

The majority of VCD devices are limited to speech which allows for NLP of the words spoken and analysis of the voice characteristics. This may be achieved by

processing the harvested data in the Cloud via a combination of "Big Data" sets, ML algorithms and the vast Compute Power resources available as depicted in Figure 4.

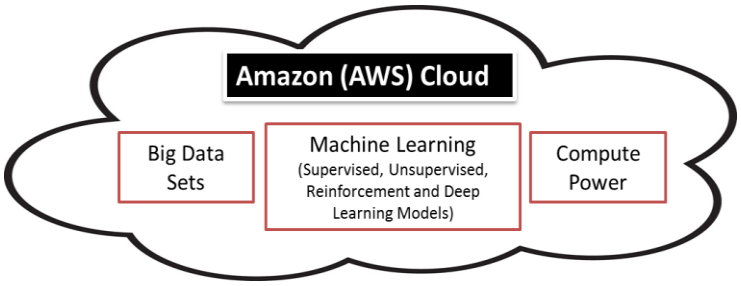

Fig. 4. Big Data Analytics in the Cloud

By enabling VCDs to act as hub for IoT devices and a myriad of applications, the manufacturers have gained access to a rich tapestry of data that relates to an individual's emotional state. Once collated and analysed, this data enables emotional state inference on a scale heretofore unknown. Furthermore, the data available to these data hungry organisations is increasing exponentially with the increased popularity of these voice command and IoT devices.

The parent companies such as Amazon who manufacture these devices are primarily 'sales organisations' and it is in their interests to market effectively. A good sales person knows their customer and the more they know, the higher the probability of making a sale [22]. Awareness of a potential customer's emotional state is extremely useful to many different types of organisations. Targeted advertising and the associated psychographic models are yielding significant returns on investment. The addition of "EMOTOGRAPHIC" models could further increase the selling power of these companies as shown in Figure 5.

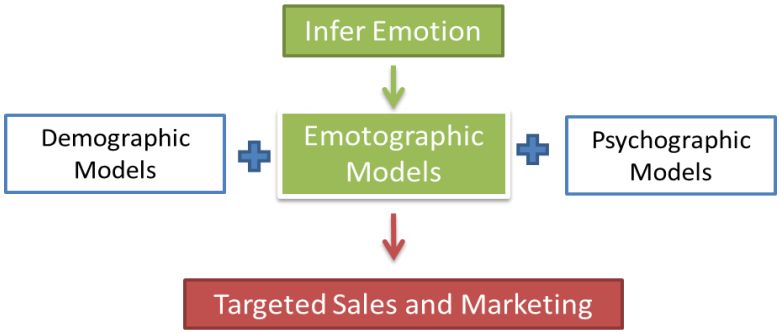

Fig. 5. Process by which Amazon may use Emotion

2.4 Detecting the Emotion of Sadness

Emotional state and mental health have become a key focus for health organisations, governments and various other organisations in the last decade [23]. Traditionally individuals suffering from afflictions such as depression will consult a health practitioner to have their mental health assessed to detect their emotional state. Mojtabai [24] describes that this is typically achieved by way of standard questions such as 'How well are you sleeping?' (MADRS), 'How are your energy levels?' (Beck Depression Inventory) or 'Have you had any thoughts of suicide?' (HAMD). Figure 6 depicts some of the inferential indictors of the emotion of sadness which also tend to be associated with the condition of depression [24]. This type of data is both interesting and valuable to organisations such as governments, pharmaceutical companies and marketing groups.

Fig. 6. Inferential Indicators of Sadness

Currently verbal responses from VCDs like Amazon's Echo do not give an indication that the devices understand emotion in humans. However, when analysis is performed on the intonation of verbal communication and data gathered from searches and linked devices such as Fitbit or music selected through streaming services, patterns may be detected that give insight to the emotional state of the user. Despite the unemotional nature of the device itself, this type of inferential data may still be collected by parent companies for use in the development of 'Affective' devices. Additionally. 'state-of-mind' has the capacity to greatly impact the needs, behaviour and habits of individuals. No matter how arbitrary, this type of data is valuable and sold to other various entities for purposes such as targeted marketing.

3 Methodology

The target VCD was an Amazon Echo that had been configured for a domestic environment and was interacted with by a single primary user over a period of five months. This device has also been configured to link to a range of external applications, as reported in [6]. Data gathered by the external applications has the potential to be utilised as a form of non-verbal information which may enable Alexa or the parent company Amazon to infer the emotional state of the user.

For the purposes of this paper it was decided to focus on sadness, one of the most common emotions that potentially may lead to an active condition of depression. To elicit information which would indicate potential levels of depression, a short experiment was conducted. A script was developed which outlined queries to be posed to the Amazon Echo VCD. These questions were designed to:

a) Establish Alexa's current 'knowledge' relating to the primary users' levels of sadness and depression. This was done in order to document verbal responses that would be offered to the same type of direct questions that would be posed by a health practitioner.

b) Establish if inferential data collected from connected devices could potentially give an indication of the primary user's emotional state.

Table 1. Devices & Applications Linked for Testing

Test Environment
Devices
• Amazon Echo 2nd Generation VCD • Fitbit Surge Wearable Health Tracker • Samsung Galaxy S5 Smart Phone • Standard Wireless Modem (Broadband Connection)
User Accounts
• Amazon Account (online shopping) • Amazon Music Account (streaming music) • Audible Account (audio books) • Fitbit Account
Applications
• Fitbit • Microsoft Office 365 (calendar/email) • Alexa Shopping List and Reminders

A test environment was constructed where an Amazon Echo VCD with default configuration was linked to several accounts, applications and IoT devices through the standard settings. It should be noted that an Amazon account is a mandatory requirement for configuring the Amazon Echo. Table 1 lists the devices, accounts and applications that were linked to the VCD for testing purposes.

4 Results & Discussion

This section offers an excerpt from the query script and responses in Table 2. In addition, a summary of the information gained is included. Each example lists the question that would be posed by a healthcare professional, the equivalent question posed to Alexa to return the desired information and also the VCD's response. Information gained directly from the Alexa application is also shown in Table 2.

Table 2. Excerpt from Query Script

Excerpt from Query Script	
Q1	Do you prefer to stay at home rather than going out and doing new things? (Geriatric Depression Scale) **Alexa, what was on my calendar on May 26th?** *Here are the four events for May 26, at 9.30 am meeting with Mary for a walk, 10.30 am meeting with Thomas, 6pm Spin Class, 8.30 drink with Tom*
Q2	How is your sleep? (MADRS) **Alexa, ask FitBit how did I sleep** You fell asleep at 11.30 pm and slept for 6hrs 45mins
Q3	How is your energy (Beck Depression Inventory) **Alexa, ask FitBit how much exercise I am taking?** *Your weekly step count is 95,000 steps, Your resting Heart rate is 53 beats per minute*
Excerpt - Amazon Alexa Application Queries	
Q4	Have you had any thoughts of suicide? (HAMD) **Search History:** *The meaning of life*
Q5	In the past two weeks how often have you felt down, depressed or hopeless (REF PHQ9) **Recent Music Searches** *REM* *The Doors*

The Alexa application documents a history of all interactions with the Echo device. The application also documents a list of all the music selections played via Amazon music. When linked with a wearable fitness tracker such as Fitbit, the device also relays information relating to the sleep patterns and activity/energy levels of the primary user. Although subtle, the combined inferential information gathered from these devices has the potential to indicate the emotional state of the user. This may be achieved though application of natural language processing algorithms to extract the key words that relate to emotion. Through these mechanisms there is the facility for the parent company, in this case Amazon, to harvest, analyses and potentially profit from the data collected.

This paper illustrates the ease by which indicators of emotions such as sadness, and indeed the associated condition of depression, may be elicited from an individual Amazon Echo device. While the probable value of this observation in terms of improving health and providing emotional support are clear, for parent companies such as Amazon, more lucrative applications of the data exist. This is inclusive of more effective and accurately targeted advertisements and recommendations. Ethically this is questionable, as the implications of this are significant when based on patterns detected relating to user emotion and potential mental health.

Amazon is primarily a retailer and the more effective the company is at marketing through target advertising, the more profitable they will become. Algorithms that indicate whether an individual may be suffering from depression could potentially lead to an unethical advertisement for medication, exercise programmes, dietary advice and counselling services. While these targeted advertisements may be helpful to the individual, companies like Amazon will increase revenue based on the unethically gathered inferential data relating to consumers.

5 Conclusion

Although emotional recognition and digital empathy are idealistic goals within the development of AI and intelligent agents, the current technology available lacks outward emotional understanding and the personas contribute only alexithymic responses. Despite extensive research by large multinational technological organizations, authentic human-like empathic interactions with intelligent agents have not yet been achieved. Consequently, users are lulled into a false sense of security where they believe that their emotions remain private.

This paper has demonstrated that despite Alexa's demonstrated lack of emotion and emotional understanding, VCDs such as the Amazon Echo have the ability to deduce emotions such as sadness through inferential data. This is displayed through responses to questions that offer the same information as those posed by health practitioners to establish potential cases of depression. This type of data paves the way for parent companies to effectively target future advertising and build EMOTOgraphic models. As users are presented with no indication of this by intelligent agents, most would be unaware that combined inferential data could be so revealing and potentially extremely profitable from a sales and marketing perspective.

As VCDs incorporate improved implementations of affective computing, the ability to build user trust as well as communicate an understanding of emotions will only increase. This potentially leads to great ethical and privacy concerns as intelligent agents such as Alexa are gradually and incrementally cured of Alexithymia indicators.

References

1. Guillotel, P., Koenig, A., Chanel, G. and Novak, D., “Toward Commercial Applications Of Affective Computing” (2015).
2. Rizzo, A., Shilling, R., Forbell, E., Scherer, S., Gratch, J. and Morency, L.P., 2016. Autonomous virtual human agents for healthcare information support and clinical interviewing. In Artificial intelligence in behavioral and mental health care (pp. 53-79).
3. Gatta, M., Dal Santo, F., Rago, A., Spoto, A. and Battistella, P.A., 2016. Alexithymia, impulsiveness, and psychopathology in nonsuicidal self-injured adolescents. Neuropsychiatric disease and treatment, 12, p.2307
4. Luminet, O., Bagby, R.M. and Taylor, G.J. eds., 2018. Alexithymia: Advances in Research, Theory, and Clinical Practice. Cambridge University Press.
5. Weik von Mossner, A., 2017. Affective Ecologies: Empathy, Emotion, and Environmental Narrative. The Ohio State University Press.
6. Furey, E., Blue, J., “She Knows Too Much – Voice Command Devices and Privacy”, 29th Irish Signals and Systems Conference (ISSC), 2018.
7. López G., Quesada L., Guerrero L.A. “Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces”. In: Nunes I. (eds) Advances in Human Factors and Systems Interaction. AHFE 2017. Advances in Intelligent Systems and Computing, vol 592. Springer, Cham, 2018.
8. Orr, D.A. and Sanchez, L., 2018. Alexa, did you get that? Determining the evidentiary value of data stored by the Amazon® Echo. Digital Investigation, 24, pp.72-78.
9. Soujanya, P., Cambria, E., Bajpai, R., Hussain, A., "A review of affective computing: From unimodal analysis to multimodal fusion." Information Fusion 37 (2017): 98-125.
10. Sokolov, D. and Patkin, M., 2018, May. Real-time emotion recognition on mobile devices. In Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on (pp. 787-787). IEEE.
11. Goksel Canbek, N., Mutlu, M., “On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants”, Journal of Human Sciences, vol 13, no1, 2016.
12. Trevarthen, C., 2017. The function of emotions in early infant communication and development. In New perspectives in early communicative development (pp. 48-81). Routledge.
13. Gulewitsch, M.D., Jusyte, A., Mazurak, N., Weimer, K. and Schönenberg, M., 2017. Preliminary evidence for increased parasympathetic activity during social inclusion and exclusion in adolescents with functional abdominal pain. Journal of psychosomatic research, 98, pp.106-112.
14. Izard, C. "Emotion theory and research: Highlights, unanswered questions, and emerging issues." Annual review of psychology 60 (2009): 1-25.
15. Yan, J., Zheng, W., Cui, Z., Tang, C., Zhang, T. and Zong, Y., 2018. Multi-cue fusion for emotion recognition in the wild. Neurocomputing.
16. Biondi, G., Franzoni, V. and Poggioni, V., 2017, July. A Deep Learning Semantic Approach to Emotion Recognition Using the IBM Watson Bluemix Alchemy Language. In International Conference on Computational Science and Its Applications(pp. 718-729). Springer, Cham.
17. Jiang, R., Ho, A.T., Cheheb, I., Al-Maadeed, N., Al-Maadeed, S. and Bouridane, A., 2017. Emotion recognition from scrambled facial images via many graph embedding. Pattern Recognition, 67, pp.245-251.

18. Stephens-Fripp, B., Naghdy, F., Stirling, D. and Naghdy, G., 2017. Automatic Affect Perception Based on Body Gait and Posture: A Survey. International Journal of Social Robotics, 9(5), pp.617-641.
19. Djara, T., Ousmane, A.M. and Vianou, A., 2018. Emotional State Recognition Using Facial Expression, Voice, and Physiological Signal. International Journal of Robotics Applications and Technologies (IJRAT), 6(1), pp.1-20.
20. Pudane, M., Lavendelis, E. and Radin, M.A., 2017. Human emotional behavior simulation in intelligent agents: processes and architecture. Procedia Computer Science, 104, pp.517-524.
21. Castellanos, S., Rodríguez, L.F., Castro, L.A. and Gutierrez-Garcia, J.O., 2018. A computational model of emotion assessment influenced by cognition in autonomous agents. Biologically Inspired Cognitive Architectures, 25, pp.26-36.
22. Cartwright, A. and Solloway, A., 2017. Emotional Intelligence: Activities for developing you and your business. Routledge.
23. Bowling, Ann. Research methods in health: investigating health and health services. McGraw-Hill Education (UK), 2014.
24. Mojtabai, R., "Diagnosing depression in older adults in primary care." New England Journal of Medicine 370, no. 13 (2014): 1180-1182.