Artificial Empathy for Clinical Companion Robots with Privacy-By-Design

Miguel Vargas Martin[1(✉)], Eduardo Pérez Valle[2], and Sheri Horsburgh[3]

[1] Ontario Tech University, Oshawa, Canada
miguel.martin@ontariotechu.ca

[2] Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Culiacán, Mexico
edupv10@hotmail.com

[3] Ontario Shores Centre for Mental Health Sciences, Whitby, Canada
horsburghs@ontarioshores.ca

Abstract. We present a prototype whereby we enabled a humanoid robot to be used to assist mental health patients and their families. Our approach removes the need for Cloud-based automatic speech recognition systems to address healthcare privacy expectations. Furthermore, we describe how the robot could be used in a mental health facility by giving directions from patient selection to metrics for evaluation. Our overarching goal is to make the robot interaction as natural as possible to the point where the robot can develop artificial empathy for the human companion through the interpretation of vocals and facial expressions to infer emotions.

Keywords: Companion robots · Mental health · Privacy-by-design · Automatic speech recognition · Artificial intelligence · Artificial empathy

1 Introduction

This paper outlines a prototype and methodology to enable a commodity humanoid robot for use as a non-pharmacological intervention to support care of individuals with dementia by enhancing the robot with privacy-by-design applications. Our use case utilizes the ASUS Zenbo (see Fig. 1), an Android-based humanoid robot with a number of built-in artificial intelligence (AI) functions that rely on the Cloud by default.

The proposed robot enhancements include the following:

1. Enhance Zenbo with privacy-enabled face expression sensing capabilities to recognize human emotions of dementia patients, off-line.

E. Pérez Valle worked on this project while at Ontario Tech, supported by a MITACS Globalink Research Internship.

J. Ye et al. (Eds.): MobiHealth 2020, LNICST 362, pp. 351–361, 2021.
https://doi.org/10.1007/978-3-030-70569-5_23

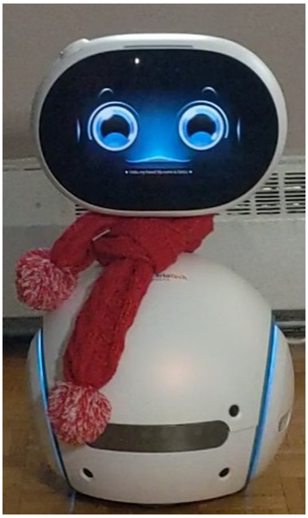


Fig. 1. ASUS' Zenbo.

2. Enhance Zenbo with privacy-enabled vocals sensing capabilities to recognize human emotions of dementia patients, off-line.
3. Identify other potential privacy-enabled functions that can be performed by Zenbo to address the needs of dementia patients and their families, such as programming individualized messages, provide lighting changes to promote calm and safe atmosphere, music therapy, reminiscence therapy and motion tracking to assist in monitoring safety of individuals.

Thus, we propose to enable Zenbo with privacy-preserving capabilities to infer human emotions by combining facial expression and voice vocals using deep learning techniques of AI through the means of specialized automatic speech recognition (ASR) hardware such as Snips [1]. In other words, we are proposing to address the area of interest involving the use of AI companions for patients and family members through systems that learn and adapt based on interactions with a situation through speech, gestures, and physical and physiological measures, amongst others. In our proposed enhanced robot, Zenbo will be able to respond when particular emotions are inferred such as sadness, anxiety, anger, etc., and will act according to clinical guidelines. For example, Zenbo may be programmed to offer the patient to show photos of the last vacation with his or her kids to make them feel better, or provide lighting changes to promote calm and safe atmosphere, offer music therapy, or even engage in a conversation with the patients.

Contributions. Our contributions are twofold: (1) We provide the technical details of our prototype privacy-by-design enhancement of ASR in Zenbo. (2) Furthermore, we describe a proposal to use a private-by-design robot to assist mental health patients and their families within the context of an inpatient mental health facility.

2 Related Work

The use of acoustic clues has been used to identify anxiety manifestations in patients with dementia (see e.g., [2]). Human-AI interaction is a growing field of research due to a persistent uncertainty about AI capabilities, and the complexity of AI's output, among other factors [3,4]. Research indicating potential benefits of assistive robots includes improving mood, communication and stress reduction [5–7]. However, many of these studies have focused on the use of robotic pets and research and functionality of humanoid robots is more limited. In a controlled clinical study [8], therapy using a humanoid robot showed a significant reduction in apathy of patients with dementia suggesting further research and development of these devices is likely of benefit to this population. Humanoid robots can include AI functions that are anticipated to provide additional support beyond those of robotic pets.

One of these robots, the Android based ASUS Zenbo is a 62 cm tall, round white body on concealed wheels, long metallic neck, with a 10″ touch screen face, no extremities, and rechargeable battery. Zenbo is a commercial humanoid robot launched in 2016 in Taiwan, marketed as a companion robot. Zenbo runs the Android operating system, and ships with a number of interactive interfaces including camera, microphone, speakers, touch sensor, drop sensor to avoid falls, range sensor to measure forward distance, and an ultrasonic sensor to avoid obstacles. The Zenbo robot has been alluded in a number of healthcare initiatives in Ontario (see e.g., [9,10]).

The factory settings of Zenbo include a number of apps, some of the most interesting ones relying on the Cloud for AI functions such as ASR. Running Android, Zenbo can be programmed through apps to perform additional functions with great flexibility by using Cloud services. Unfortunately, there has been a number of concerns regarding privacy of Cloud services, and devices like Amazon's Alexa or Google Home aren't the exception, with a number of privacy breach incidents making the news [11–14].

Privacy-preserving hardware like Sonos' Snips [15], a powerful Raspberry Pi card shipping with built-in specialized ASR software, provide a solution to privacy concerns of ASR Cloud services as they are capable of performing all the computations off-line, removing the need for the Cloud. Snips can thus be trained with intents to create voice assistants that understand spoken languages (English, French, Spanish, German, Italian, Japanese, and Korean).

Finally, we note that ASR systems are susceptible to a number of attacks that can seriously hinder the technology. For example, it has been demonstrated how hidden voice commands can be effectively interpreted by voice recognition systems while being imperceptible by humans [16,17]. A hidden voice attack, then, may trigger unwanted and unexpected robot behaviour which could be found unacceptable within certain contexts.

3 Prototype Private-By-Design ASR in Zenbo

The prototype in our Human Machine Laboratory uses Snips, which offers private-by-design ASR, to listen and process voice commands and send them to Zenbo for execution. The prototype includes commands such as “tell me a joke” or “how’s the weather?” or “make a happy face”. To achieve this, it was necessary to (1) configure Snips to interpret voice commands and create a secure connection with Zenbo, and (2) develop an Android app within Zenbo which can interpret the commands received from Snips.

3.1 Configuring Snips

There are several platforms where Snips can be installed; our prototype used Raspberry Pi. The configuration steps can be found in the Snips documentation [18]. Once Snips is configured for Raspberry Pi, we use the Snips Console to create different types of assistants which will allow us to create a connection with Zenbo. Figure 2 depicts an assistant called *HelloSnips*. Then we can create apps inside the assistant; we called our app *Zenbo*, as illustrated in Fig. 3.

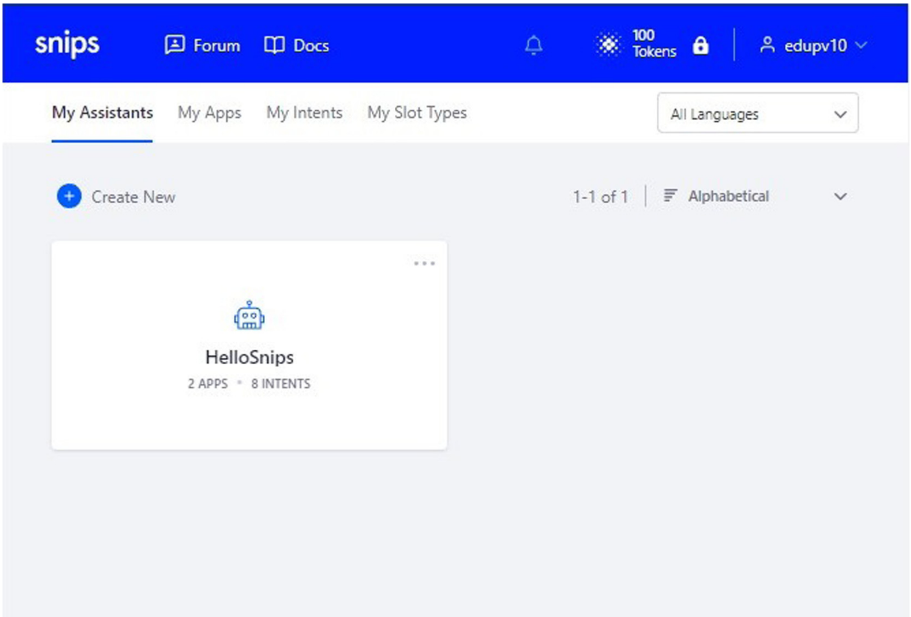
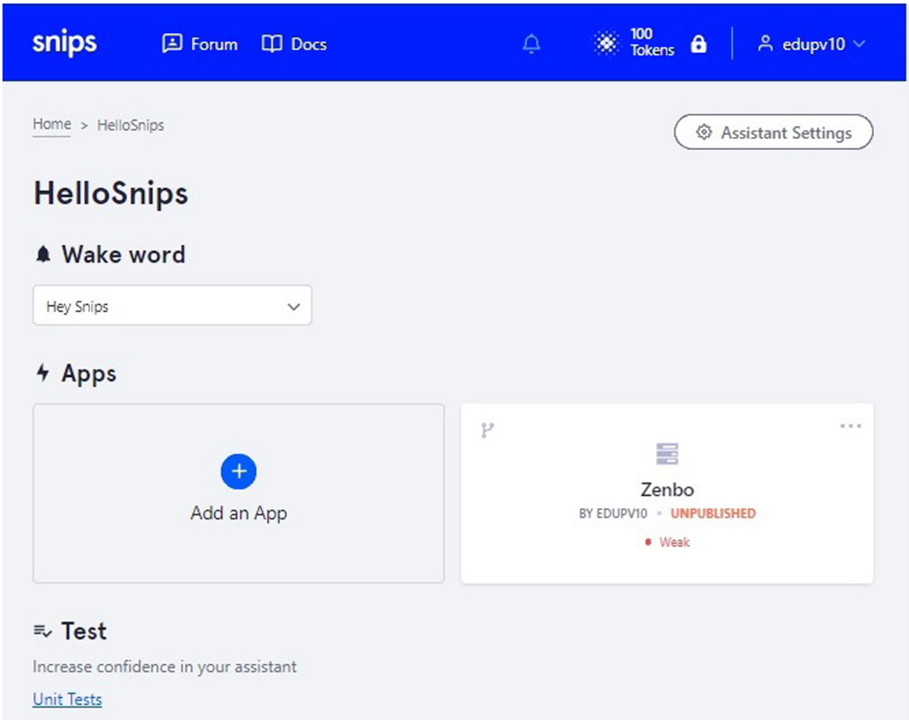


Fig. 2. The assistant shown is called *HelloSnips*.



HelloSnips

Wake word

Hey Snips

Apps

Zenbo

Test

Increase confidence in your assistant

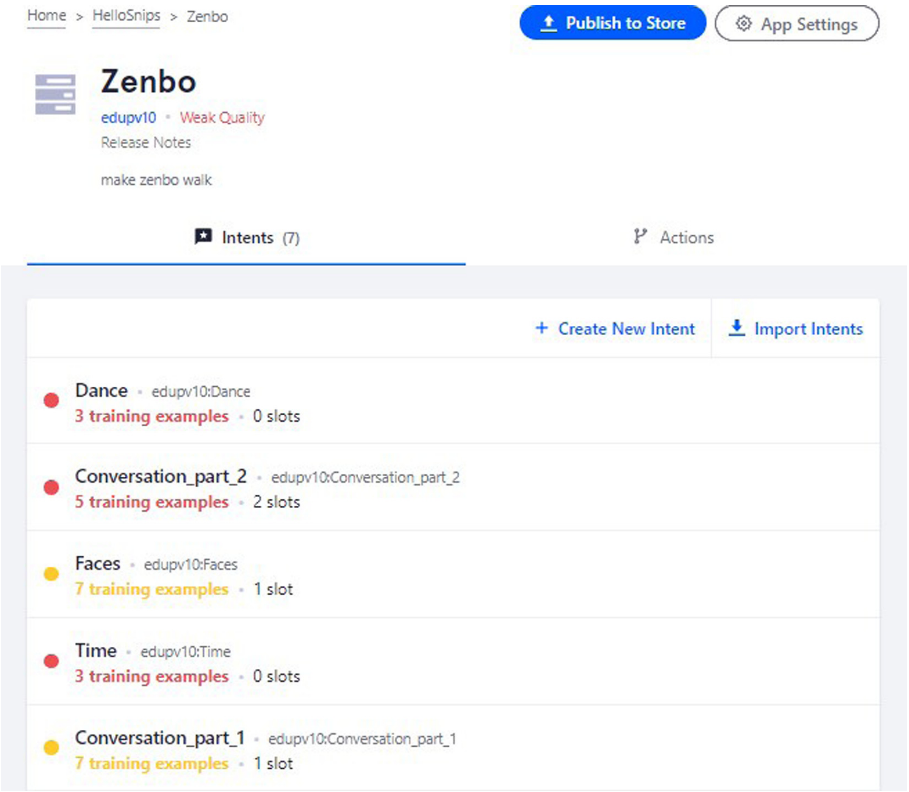
Fig. 3. An application called *Zenbo* within an assistant.

The applications for the Snips assistant are composed of intents and actions, where for every action there can be one or more intents. Intents are the sentences that Snips recognizes and can trigger some action on Zenbo. Figure 4 illustrates some intents.

Within the intent *Faces* we used 7 examples, as shown in Figure 5. So, when Snips identifies the intent it will send the appropriate command to Zenbo to trigger the corresponding action on the robot. The words marked in blue are called *Slots*, which are benchmarks used to recognize intents and link them to the corresponding action. Actions are coded in Python inside *Code Snippets*, as illustrated in Fig. 6. In our prototype the action is the establishment of a secure communication channel with Zenbo where commands are then sent to the robot.

3.2 Android App in Zenbo

In our prototype, Zenbo receives commands via sockets through the WiFi network. The sockets are established with Zenbo as a server, and Snips as a client. To develop the Android app in Zenbo, we used Zenbo's SDK available at [19]. Once the app is built, it can be seamlessly installed in Zenbo. We leave the underpinnings of the app outside the scope of this paper but suffice to say that the app receives the text of the intent from Snips, and then cross-check it against a set of possible actions (Zenbo commands); and execute the one that matches the intent.



Zenbo

edupv10 · Weak Quality

Release Notes

make zenbo walk

● Dance · edupv10:Dance
3 training examples · 0 slots

● Conversation_part_2 · edupv10:Conversation_part_2
5 training examples · 2 slots

● Faces · edupv10:Faces
7 training examples · 1 slot

● Time · edupv10:Time
3 training examples · 0 slots

● Conversation_part_1 · edupv10:Conversation_part_1
7 training examples · 1 slot

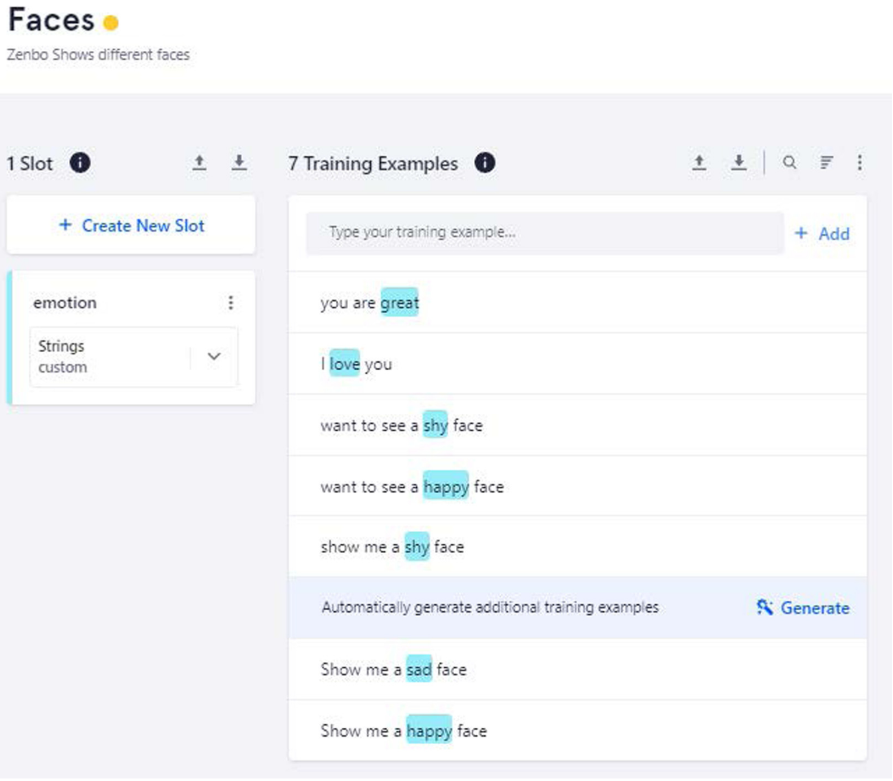
Fig. 4. Examples of intents used in our prototype.

4 Proposed Enhancements and Methodology for Use in a Clinical Environment

The prototype described in the previous section proves the viability of a private-by-design ASR robot. Now we describe two constructs that will make Zenbo a feasible clinical robot.

4.1 Technical Aspects

One construct involves the technical aspects of training a sufficiently large number of intents for every action in such a way that Zenbo can understand a large range of voice commands. Another aspect of this construct is related to vocals and face expression recognition (FER). The second construct is the effectiveness in identifying dementia-specific needs and being able to address these with our technology, and having our solution tested by real patients.



Faces

Zenbo Shows different faces

1 Slot

7 Training Examples

emotion

Strings custom

you are great

I love you

want to see a shy face

want to see a happy face

show me a shy face

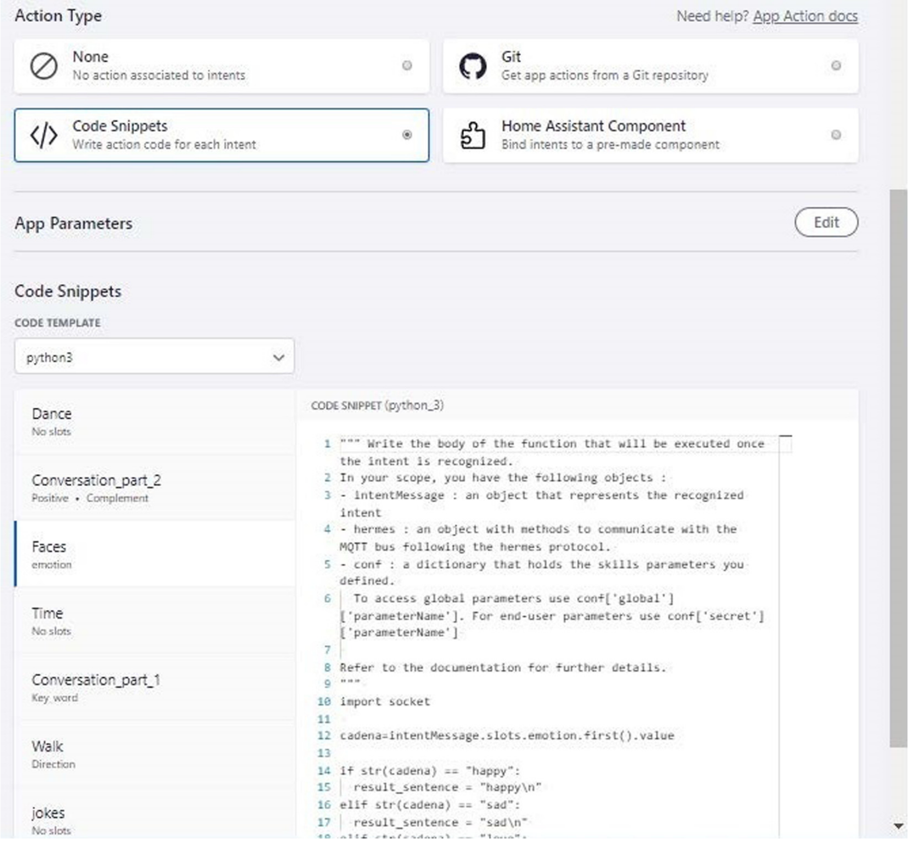
Automatically generate additional training examples

Show me a sad face

Show me a happy face

Fig. 5. Some training examples within the *Faces* intent.

Vocals and Face Expression Recognition. It's been generally accepted in the research community that acoustic profiles of vocals are associated to emotions of anger, fear, happiness, and sadness. The profiling of vocals includes pitch, intensity, and speed of speech [20]. We will use deep learning to train a vocals engine to recognize these emotions. And although these vocal profiles are generally accepted, we are aware that emotions may manifest differently on individuals, so we will further consider the possibility to customize and refine the training with the vocals of the dementia patient the robot will serve. As per FER, different techniques have been proposed in the literature to detect expressions such as smile, sad, anger, disgust, surprise, and fear [21]. While FER feature extraction is one of the most difficult challenges of our project, Zenbo comes with pre-installed basic face recognition features that we hope can be adapted for our purposes. We plan to complete the vocals recognition and combine that with FER to infer emotions of the dementia patient. This, combined with the NLP engine in a privacy-enhanced environment will make of our approach a powerful tool with great potential to assist dementia patients.



CODE TEMPLATE: python3

CODE SNIPPET (python_3)

```
""" Write the body of the function that will be executed once the intent is recognized.
In your scope, you have the following objects :
- intentMessage : an object that represents the recognized intent
- hermes : an object with methods to communicate with the MQTT bus following the hermes protocol.
- conf : a dictionary that holds the skills parameters you defined.
  To access global parameters use conf['global']['parameterName']. For end-user parameters use conf['secret']['parameterName']

Refer to the documentation for further details.
"""
import socket

cadena=intentMessage.slots.emotion.first().value

if str(cadena) == "happy":
  result_sentence = "happy\n"
elif str(cadena) == "sad":
  result_sentence = "sad\n"
```

Fig. 6. Snips code template to develop the commands to execute corresponding actions.

Emotions. We will use a dimensional theory of emotion based on the "PAD theory" [22] for pleasure (a measure of valence), arousal (a measure of affective activation), and dominance (a measure of control), termed the valence-arousal model. This model classifies emotions such as sad, happy, and calm, and is able to associate intensities to these [23]. We will be particularly interested in detecting emotions that are deemed as requiring attention of a human caregiver. For example, after using vocals and FER, our system could trigger an alert to the floor nurse if unusual sadness emotions are detected. In addition, our system will be able to pick significantly different vocal features that may be an indication of pain or any other condition such as skipping medications, or sudden changes in the patients' condition.

4.2 Clinical Trials

The second construct of this proposal is the effective identification of dementia patients needs that can be addressed with the physical limitations of the Zenbo robot.

Test Group. The ideal test group would be patients with a primary diagnosis of dementia and behaviours and psychological symptoms of dementia (BPSD). BPSD can include agitation, aggression, restlessness, lability, exit seeking, impulsivity and sexual disinhibition. A trigger for BPSD can be boredom or social isolation, therefore supporting activities and increasing non-pharmacological interventions is critical to stabilizing behaviours in this population. Common non-pharmacological interventions include the use of robotic pets, doll therapy, aromatherapy, therapeutic sensory chair, sensory room (Snoezelen room), SPA-based therapy (e.g., manicure), iPads/iPods, live pet therapy, hand massage, and group activities.

Patient Identification. To identify patient needs, we will look to partner with facilities caring for individuals with dementia, conduct focus groups with frontline staff, as well as individual patients and/or their caregivers. Themes from these focus groups will be reviewed by the research team for exploration of the capabilities of Zenbo to support these activities. Upon enhancement of the Zenbo robot, we will look to pilot the device with a number of dementia patients, determined by the Test Group assessment described above.

Evaluation. To evaluate the effectiveness of the intervention, baseline measures for each patient including DSM-V (Diagnostic and Statistical Manual of Mental Disorders' Working Group 5) diagnoses, Folstein Mini-Mental state Examination (MMSE) on admission, Alzheimer's Disease-related Quality of Life (QoL-AD) scale on admission, baseline measures on the neuropsychiatric inventory, number of incidents of aggression/threatening/sexually inappropriate behaviour and average number of hours slept per night during the week prior to the intervention will be reviewed. Outcome measures will include neuropsychiatric inventory, number of incidents of aggression/threatening/sexually inappropriate behaviour, and average number of hours slept per night the week post the intervention, for comparison. In addition, the QoL-AD will be completed post-intervention as well.

5 Conclusions and Directions for Future Work

The research study will enable Zenbo with privacy-preserving capabilities to infer human emotions by combining facial expression and voice vocals using deep learning techniques of AI. When Zenbo is enabled with these capabilities they will serve as a meaningful companion for individuals with dementia, thus improving the quality of life of these individuals. This will also provide an additional non-pharmacological intervention to support stabilization of BPSD, which will be transferable across acute care, tertiary care, and community care settings (e.g., long-term care homes). In addition, confirming the privacy-preserving capabilities will allow for adaptation of this model to potentially meet the needs of individuals with dementia who continue to live at home. Given the prediction that in 20 years' time over 1.5 million Canadians will be living with dementia and the significant economic burden associated with providing meaningful support and care to these individuals, identifying cost-effective ways to support independence and quality of life will be crucial.

The Zenbo private-by-design approach also has the potential to combine several non-pharmacological interventions into one device as through this study we will be able to investigate incorporation of light therapy, music therapy, reminiscence therapy and potential safety monitoring which negates the expense and space requirements of having multiple devices to provide these interventions.

Improvements to our current prototype include training the private-by-design ASR with sufficiently large number of intents and samples per intent, which will make Zenbo recognize speech in a more natural way. To this end, we could choose to use transfer learning [24], making sure to protect against potential backdoor attacks [25]. Alternatively, the Snips device could be replaced by a fully-fledged ASR that works locally without Cloud services, although this would require expensive equipment which may affect the portability of the hardware set.

References

1. Coucke, A., et al.: Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. ArXiv abs/1805.10190 (2018)
2. Hernandez, N., et al.: Prototypical system to detect anxiety manifestations by acoustic patterns in patients with dementia. PHAT **5**(19) (2019)
3. Yang, Q., et al.: Re-examining whether, why, and how Human-AI interaction is uniquely difficult to design. In: Conference on Human Factors in Computing Systems (CHI), Honolulu, USA (2020)
4. Long, D., et al.: What is AI literacy? Competencies and design considerations. In: Conference on Human Factors in Computing Systems (CHI), Honolulu, USA (2020)
5. Murdoch, E., et al.: Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas **74**, 14–20 (2013)
6. Broekens, J., et al.: Assistive social robots in elderly care: a review. Gerontechnology **8**(2), 94–103 (2009)
7. Bemelmans, R., et al.: Socially assistive robots in elderly care: a systematic review into effects and effectiveness. JAMDA **13**(2), 114–120 (2012)
8. Soler, M.V., et al.: Social robots in advanced dementia. Front. Aging Neurosci. **7**(133), 1–12 (2015)
9. Perkins, J.: Toronto charity creates robot to entertain, educate kids who can't go to school due to severe illnesses. The Globe and Mail, 31 January 2020. https://www.theglobeandmail.com/canada/toronto/article-toronto-charity-creates-robot-to-entertain-educate-kids-who-cant-go/. Accessed 23 Nov 2020

10. Students using AI to teach robot how to recognize human emotions. CTV News, 19 August 2019, http://ctv.news/6JsxuKV. Accessed 23 Nov 2020
11. Brown, J.: The Amazon Alexa eavesdropping nightmare came true. Gizmodo. https://gizmodo.com/the-amazon-alexa-eavesdropping-nightmare-came-true-1831231490. Accessed 23 Nov 2020
12. Valinski, J.: Amazon reportedly employs thousands of people to listen to your Alexa conversations. CNN Business. https://www.cnn.com/2019/04/11/tech/amazon-alexa-listening/index.html. Accessed 23 Nov 2020
13. Paul, K.: Google workers can listen to what people say to its AI home devices. The Guardian. https://www.theguardian.com/technology/2019/jul/11/google-home-assistant-listen-recordings-users-privacy. Accessed 23 Nov 2020
14. Barack, L.: Google Home security breach sends your location to hackers. GearBrain. https://www.gearbrain.com/google-home-location-hack-found-2579276699.html. Accessed 23 Nov 2020
15. Snips: Using voice to make technology disappear. https://snips.ai/. Accessed 23 Nov 2020
16. Chen, Y., et al.: Devil's Whisper: a general approach for physical adversarial attacks against commercial black-box speech recognition devices. In: 29 USENIX Security Symposium (2020)
17. Abdullah, M., et al.: Practical hidden voice attacks against speech and speaker recognition systems. In: Network and Distributed System Security Symposium (NDSS), San Diego, USA (2019)
18. Quick Start Raspberry Pi. https://docs.snips.ai/getting-started/quick-start-raspberry-pi. Accessed 23 Nov 2020
19. ASUS Developer. https://zenbo.asus.com/developer/tools/. Accessed 23 Nov 2020
20. Juslin, P.N., et al.: Communication of emotion in vocal expression and music performance: different channels, same code? Psychol. Bull. **129**, 770–814 (2003)
21. Revina, I.M., et al.: A survey on human face expression recognition techniques. Psychol. Bull. (2018). https://doi.org/10.1016/j.jksuci.2018.09.002
22. Albert, M., et al.: An Approach to Environmental Psychology. The MIT Press, Cambridge (1974)
23. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
24. Hernandez, N., et al.: Literature review on transfer learning for human activity recognition using mobile and wearable devices with environmental technology. SN Comput. Sci. **1**, 66 (2020)
25. Yao, Y., et al.: Latent backdoor attacks on deep neural networks. In: ACM Conference on Computer and Communications Security, London, UK (2019)