Let's Talk it through, anew: Promises and Pitfalls of Customisable Conversational Reflection Support

Mateusz Dubiel
University of Luxembourg
Luxembourg
mateusz.dubiel@uni.lu

Luis A. Leiva
University of Luxembourg
Luxembourg
luis.leiva@uni.lu

ABSTRACT

As modern lifestyles are becoming increasingly stressful and ever more hectic with multiple stimuli constantly competing for our attention, Affective Disorders (ADs) such as anxiety and depression are on the rise. Consequently, due to the burgeoning demand for counselling and therapeutic services, many people who suffer from ADs are struggling to timely access the professional support that they require. To address this problem, voice-enabled Conversational Agents (CAs) have been recently proposed as tools for supporting self-reflection and providing assistance in managing a range of ADs through synthetic voices. However, despite their therapeutic potential, CAs offer a very limited choice when it comes to selection and personalisation of synthetic voices used. The goal of this paper is two-fold: (1) it discusses the potential benefits that a CA's voice customisation can bring to enhance user engagement and promote long term self-reflection, and (2) it offers reflection on the corresponding challenges associated to this approach.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; *Auditory feedback*.

KEYWORDS

self-reflection; conversational agents; behavioural change; synthetic speech

ACM Reference Format:
Mateusz Dubiel and Luis A. Leiva. 2023. Let's Talk it through, anew: Promises and Pitfalls of Customisable Conversational Reflection Support. In *CHI'23 Workshop on Integrating Individual and Social Contexts into Self-Reflection Technologies.* ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

> "Between stimulus and response there is a space.
> In that space is our power to choose our response.
> In our response lies our growth and our freedom." [16]
> —*Viktor E. Frankl*

Self-reflection has proven to be a highly effective tool in reducing stress levels [9, 37] and helping people to manage their social anxiety [3, 35]. Thanks to the proliferation of smartphones and other types of mobile devices, digital interventions are currently used to support a growing number of users in improving their well-being [10, 15, 20, 22], enhancing productivity and focus [4, 18], and making them more physically active [23, 36, 39]. However, regardless of its potential to bring about positive behavioural change, self-reflection is a highly-complex and challenging task which requires dedication and long-term commitment. According to Nahum-Shani et al., who coined the term 'Just-in-Time Adaptive Interventions' (JITAIs), an effective digital intervention needs to adapt to the ever-changing state of the user and account for the interaction context in order to successfully capture attention and elicit desirable behaviour [31]. The key factors that determine individuals' adherence to JITAIs are engagement and fatigue [31], and ensuring a proper balance between the provision of the automated support and personal volition and agency of the user [34]. Indeed, recent research shows that users value mental health applications that are straightforward to use, offer a variety of options to chose from, and allow for customisation of the functionalities provided [2].

In this position paper, building on previous research on self-reflection tools, we will discuss how the application of personalised and customisable, synthetic speech technology can help to increase adherence to digital interventions by increasing engagement, reducing fatigue, and giving users the ability to exercise their agency. We will also consider challenges posed by this approach and make some recommendations for the future research agenda for self-reflection CAs. The idea of providing users with more control over how they receive the CA's support is based on the view advocated by Mobbs et al. who consider 'free will' as a fundamental prerequisite in self-reflection and self-improvement, since people need to believe that they are in full control of their actions in order to strive for change [30].

2 SUPPORTING SELF-REFLECTION IS A CHALLENGING TASK

Research indicates that digital interventions tend to become less effective over time and ultimately lead to abandonment after just several uses [17]. In the context of eHealth applications, this phenomenon is known as 'the law of attrition' [13]. In order to reduce the likelihood of abandonment and increase adherence to digital interventions, a person needs to perceive them as useful and trustworthy [21]. What is more, since processing digital interventions requires engagement of cognitive resources, it is crucial to adjust the scope and style of the message, as well as its delivery frequency so that the user is receptive of it [14]. Otherwise, in the long term, too high frequency of messages that lack variability can lead to intervention fatigue [11], causing participants disengage from the intervention due to boredom or cognitive overload.

CHI'23 Workshop: Integrating Individual and Social Contexts into Self-Reflection Technologies, April 28, 2023, Hamburg, Germany

Given that signal intervention fatigue can be detrimental to engagement [6], it is crucial to deliver interventions in a way that is engaging. This can be achieved by varying the form of content delivery [26] and giving the user more control regarding how the supportive process is being delivered [31]. The main challenge is to determine *how* to best incorporate participant-determined features to digital interventions to balance the provision of the automated support with personal volition and agency. Here, we propose that providing more control over synthetic speech used by the self-reflection CA could help to achieve this goal.

3 POTENTIAL OF SPEECH IN SELF-REFLECTION

Research indicates that even non-interactive learning materials such as audio recordings can increase perception of social presence and, in turn, result in an improved learning performance [29]. This effect is linked to a phenomenon called 'social facilitation' [38], where performance is improved as a consequence of the felt social presence. Research also shows that feedback provided through media materials helps to establish a stronger social presence of the instructor [1]. Human voices were also identified as triggers for increased social presence [24]. Thanks to recent technological improvements, the quality of synthetic speech is currently sufficient for credible and effective delivery of verbal information [8]. This, in turn, provides an opportunity to reliably convey social cues during self-reflection and improve intervention adherence.

Indeed, recent studies (e.g., [27, 28] acknowledge the usefulness of voice-enabled CAs in supporting self-reflection. Maharjan et al. [28] found that a smart-speaker-based CA 'Sofia' was perceived as significantly more attractive and novel compared to a web-based self-reflection application that did not feature speech. While there was no significant difference in terms of user engagement, Sofia's engagement scores remained consistent over the span of four weeks while the web based app engagement decreased as the trial went on. The authors highlighted that the currently available CAs are lacking personalisation, which in the long term may lead to discontinued use, as the novelty effect fades away [28]. Mharjan et al. [27] have also found that perception of Sofia and the ways in which participants personified this CA differed between the users — with some people exhibiting a positive while others a negative sentiment towards the agent. Therefore, in order to best support a subjective users' experience, individual needs and personal differences need to be accounted for. Accordingly, self-reflection CAs need to 'provide space for human experiences of socialising, connectedness, empathy and compassion, while allowing users to appropriate technology in the ways they see fit' [27]. We believe that giving participants control over type of the voice used by their CA could facilitate the self-reflection and make it more engaging and effective.

Studies with human speech showed that enthusiastic voices are ranked higher in terms of their social and affective qualities and lead to an increased focus and better performance, as compared to calm voices [25]. When it comes to synthetic voices, it was observed that high-quality synthetic speech received higher ratings for credibility and engagement, compared to low quality synthetic speech [5]. A high-quality synthetic speech has also been shown to increase perception of trust [7]. However, it should be noted that while speech has a clear potential to motivate users, the vast majority of possible voice factors, such as tone, pitch, intonation, etc. have not been fully explored yet in terms of social processes [33]. This opens up new research avenues for evaluating the impact of synthetic speech in the context of digital interventions.

We believe that the HCI community who work on design and development of reflection interventions should consider the following challenging research gaps, in order to make the design of CA more engaging and improve the prospects of long-term adherence.

- ***How to select the voice of the agent to increase user engagement?*** Dubiel et al. [12] demonstrated that the type of speech corpus selected for voice development should be aligned with the intended purpose of its usage.
- ***Should the CA match the conversational style of the user in terms of prosodic qualities of the voice?*** Hoegen et al. [19] found that individuals with High Consideration conversational styles were more likely to trust a CA that matched their conversational style.
- ***How should we pick the personalisation criteria?*** Should personalisation criteria be determined by personality questionnaires, or automatically elicited over the course of the CA use, based on the user's behaviour?
- ***How to prevent habituation and over-reliance?*** Should users be encouraged to change the type of the voice used by CA? If so, at what point should such a recommendation be delivered?

We argue that the discussion of the above points could help in determining the future research agenda for reflection-support CAs, and provide insights if engaging and customisable speech can facilitate the long-term intervention adherence.

4 PITFALLS OF CUSTOMISABLE SPEECH SUPPORT AND THE ROAD AHEAD

Finally, we would like to acknowledge that while customisable self-reflection CAs can bring a range of benefits, there are also several pitfalls that should be considered. First, long-term exposure to CA can result in addiction and in turn harm real-life intimate relationships, as in the case of the AI-driven social chatbot Replica [40]. Second, and related to first point, as the relationship with the CA develops, users may rely on it exclusively and cease to seek a professional support from therapists and counsellors whose expertise may be necessary for the user to properly manage their AD and prevent it for deteriorating. Third, on the point of accessibility, not all users may have access to devices that support speech interaction, or due to hardware limitations can only support low-quality synthetic voices, therefore it is important to consider alternatives that can be used to bridge this digital divide.

Overall, the goal of implementing personalised synthetic speech is to facilitate self-reflection and make it more engaging and less taxing. However, it should be noted that while important, it is not considered to be a replacement for human interaction. Therefore, in addition to voice customisation, the design of voice-enabled CAs should also promote social interactions such as pointing users to relevant online communities and peer groups that could offer users further support. Consequently, this could further promote user agency giving people more choice over the intervention options [32]. The ultimate goal for making CAs more customisable is to provide

support that is user-centred rather than intervention-centered. As posited by Nahum-Shrani et al. [31] while theoretical overviews can guide developers of self-reflection tools, cross-disciplinary collaborations and users evaluations are required to ensure that the technology is appropriately used. This paper provides a speech processing perspective that should be validated through experimentation.

ACKNOWLEDGMENTS

This research is supported by the Horizon 2020 FET program of the European Union (grant CHIST-ERA-20-BCI-001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

REFERENCES

[1] Steven R Aragon. 2010. *Facilitating Learning in Online Environments: New Directions for Adult & Continuing Education, No. 100.* Vol. 103. John Wiley & Sons.
[2] Andreas Balaskas, Stephen M Schueller, Anna L Cox, Chuck Rashleigh, and Gavin Doherty. 2023. Examining young adults daily perspectives on usage of anxiety apps: A user study. *PLOS Digital Health* 2, 1 (2023), e0000185.
[3] Georg F Bauer, Oliver Hämmig, Wilmar B Schaufeli, and Toon W Taris. 2014. A critical review of the job demands-resources model: Implications for improving work and health. *Bridging occupational, organizational and public health: A transdisciplinary approach* (2014), 43–68.
[4] Jennifer R Burnett and Timothy C Lisk. 2019. The future of employee engagement: Real-time monitoring and digital tools for engaging a workforce. *International Studies of Management & Organization* 49, 1 (2019), 108–119.
[5] Erin K Chiou, Noah L Schroeder, and Scotty D Craig. 2020. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education* 146 (2020), 103756.
[6] Michael S Cole, Frank Walter, Arthur G Bedeian, and Ernest H O'Boyle. 2012. Job burnout and employee engagement: A meta-analytic examination of construct proliferation. *Journal of management* 38, 5 (2012), 1550–1581.
[7] Scotty D Craig, Erin K Chiou, and Noah L Schroeder. 2019. The impact of virtual human voice on learner trust. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 63. SAGE Publications: Los Angeles, CA, 2272–2276.
[8] Scotty D Craig and Noah L Schroeder. 2019. Text-to-speech software and learning: Investigating the relevancy of the voice effect. *Journal of Educational Computing Research* 57, 6 (2019), 1534–1548.
[9] Monique F Crane, Maria Kangas, Eyal Karin, Ben Searle, and Diana Chen. 2020. Leveraging the experience of stressors: the role of adaptive systematic self-reflection. *Anxiety, Stress, & Coping* 33, 3 (2020), 231–247.
[10] Mary Czerwinski, Javier Hernandez, and Daniel McDuff. 2021. Building an AI That Feels: AI systems with emotional intelligence could learn faster and be more helpful. *IEEE Spectrum* 58, 5 (2021), 32–38.
[11] Sara Demain, Ana-Carolina Goncalves, Carlos Areia, Ruben Oliveira, Ana Jorge Marcos, Alda Marques, Ranj Parmar, and Katherine Hunt. 2015. Living with, managing and minimising treatment burden in long term conditions: a systematic review of qualitative research. *PloS one* 10, 5 (2015), e0125457.
[12] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.
[13] Gunther Eysenbach et al. 2005. The law of attrition. *Journal of medical Internet research* 7, 1 (2005), e402.
[14] Joel E Fischer, Nick Yee, Victoria Bellotti, Nathan Good, Steve Benford, and Chris Greenhalgh. 2010. Effects of content and time of delivery on receptivity to mobile interruptions. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 103–112.
[15] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e7785.
[16] Viktor E Frankl. 1985. *Man's search for meaning*. Simon and Schuster.
[17] Yoshimi Fukuoka, Caryl Gay, William Haskell, Shoshana Arai, Eric Vittinghoff, et al. 2015. Identifying factors associated with dropout during prerandomization run-in period from an mHealth physical activity education study: the mPED trial. *JMIR mHealth and uHealth* 3, 2 (2015), e3928.
[18] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and evaluation of intelligent agent prototypes for assistance with focus and productivity at work. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 390–400.
[19] Rens Hoegen, Deepali Aneja, Daniel McDuff, and Mary Czerwinski. 2019. An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 111–118.
[20] Becky Inkster, Shubhankar Sarda, Vinod Subramanian, et al. 2018. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth* 6, 11 (2018), e12106.
[21] Eline C Jochems, Cornelis L Mulder, Arno van Dam, Hugo J Duivenvoorden, Sylvia CM Scheffer, Willem van der Spek, and Christina M van der Feltz-Cornelis. 2012. Motivation and treatment engagement intervention trial (MotivaTe-IT): The effects of motivation feedback to clinicians on treatment engagement in patients with severe mental illness. *BMC psychiatry* 12 (2012), 1–17.
[22] Rafal Kocielnik, Raina Langevin, James S George, Shota Akenaga, Amelia Wang, Darwin P Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T Hsieh, et al. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–10.
[23] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: a conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–26.
[24] Eun-Ju Lee. 2010. What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research* 37, 2 (2010), 191–214.
[25] Tze Wei Liew, Su-Mae Tan, Teck Ming Tan, and Si Na Kew. 2020. Does speaker's voice enthusiasm affect social cue, cognitive load and transfer in multimedia learning? *Information and Learning Sciences* (2020).
[26] Fred Mael and Steve Jex. 2015. Workplace boredom: An integrative model of traditional and contemporary approaches. *Group & Organization Management* 40, 2 (2015), 131–159.
[27] Raju Maharjan, Kevin Doherty, Darius Adam Rohani, Per Bækgaard, and Jakob E Bardram. 2022. Experiences of a Speech-enabled Conversational Agent for the Self-report of Well-being among People Living with Affective Disorders: An In-the-Wild Study. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 2 (2022), 1–29.
[28] Raju Maharjan, Darius A Rohani, Kevin Doherty, Per Bækgaard, and Jakob E Bardram. 2022. What Is the Difference? Investigating the Self-Report of Wellbeing via CA and Web App. *IEEE Pervasive Computing* 21, 2 (2022), 60–68.
[29] Richard E Mayer. 2014. Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. *The Cambridge handbook of multimedia learning* 16 (2014), 345–370.
[30] Dean Mobbs, Hakwan C Lau, Owen D Jones, and Chris D Frith. 2009. *Law, responsibility, and the brain*. Springer.
[31] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.
[32] Herman Saksono, Carmen Castaneda-Sceppa, Jessica A Hoffman, Magy Seif El-Nasr, and Andrea Parker. 2021. StoryMap: Using Social Modeling and Self-Modeling to Support Physical Activity Among Families of Low-SES Backgrounds. In *CHI 2021 Conference on Human Factors in Computing Systems*. 1–14.
[33] Sascha Schneider, Maik Beege, Steve Nebel, Lenka Schnaubert, and Günter Daniel Rey. 2021. The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review* (2021), 1–38.
[34] Stephen M Schueller, Kathryn Noth Tomasino, and David C Mohr. 2017. Integrating human support into behavioral intervention technologies: The efficiency model of support. *Clinical Psychology: Science and Practice* 24, 1 (2017), 27.
[35] Catrinel A Stefan and Lavinia Cheie. 2022. Self-compassion and social anxiety in late adolescence: Contributions of self-reflection and insight. *Self and Identity* 21, 2 (2022), 210–222.
[36] Stephanie Stockwell, Patricia Schofield, Abi Fisher, Joseph Firth, Sarah E Jackson, Brendon Stubbs, and Lee Smith. 2019. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: A syst. review and meta-analysis. *Experimental Gerontology* 120 (2019), 68–87.
[37] Cheryl J Travers, Dominique Morisano, and Edwin A Locke. 2015. Self-reflection, growth goals, and academic outcomes: A qualitative study. *British journal of educational psychology* 85, 2 (2015), 224–241.
[38] Norman Triplett. 1898. The dynamogenic factors in pacemaking and competition. *The American journal of psychology* 9, 4 (1898), 507–533.
[39] Saskia Van Dantzig, Gijs Geleijnse, and Aart Tijmen Van Halteren. 2013. Toward a persuasive mobile application to reduce sedentary behavior. *Personal and ubiquitous computing* 17 (2013), 1237–1246.
[40] Tianling Xie and Iryna Pentina. 2022. Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika. (2022).