Artificial Intelligence Supported Cognitive Behavioral Therapy for Treatment of Speech Anxiety in Virtual Reality Environments

Fredrik Åhs, Department of Psychology and Social Work, Mid Sweden University, Sweden

Peter Mozelius, Department of Computer and System Science, Mid Sweden University, Sweden

Felix Dobslaw, Department of Computer and System Science, Mid Sweden University, Sweden

Abstract: Cognitive behavioral therapy (CBT) has become a successful treatment to improve management of stress and anxiety in social situations. One of the most widespread social anxiety disorders is speech anxiety, and there are also studies reporting that speech anxiety is increasing among younger adults. An emerging trend in CBT treatment is virtual reality (VR), a technology that today also could involve the use of artificial intelligence. The aim of this position paper is to present and discuss the idea of using explainable artificial intelligence to improve CBT treatment of speech anxiety in virtual reality environments.

The proposed CBT and VR concept builds upon identification of individuals for whom a scientifically grounded treatment can be predicted to have a larger effect than the average. The identification of these individuals should be conducted with the use of Explainable artificial intelligence (XAI). However, the effect of providing XAI-based information on actual treatment outcome has not been fully investigated and established. To better understand how AI-based information can strengthen CBT, it would be valuable to investigate how much confidence individuals undergoing treatment can have in information that is derived from XAI applications. If XAI-derived information is trusted to the same extent as traditional information coming from psychologists, this could open up for CBT design.

Furthermore, the VR-treatment should be grounded in learning theory and cognitive psychology with an emphasis on promotion of inhibitory learning. A commercial application should be used for stimuli presentation in the VR-head-set based on various scenarios that simulates real-world situations. The main objective of the VR-treatment is to promote inhibitory learning by disproving catastrophic beliefs through exposure to distressful speech situations. Outcomes of the treatment should primarily be measured by the Public Speaking Anxiety Scale, but also involve an assessment of social anxiety with the use of Liebowitz's Social Anxiety Scale.

Keywords: Artificial intelligence, Explainable artificial intelligence, Cognitive behavioral therapy, Virtual reality, Speech anxiety

Introduction and aim

Speech anxiety can be seen as the hallmark of social anxiety disorder (SAD), and something that involves regular problems. The prevalence in Sweden is relatively high with 15.6% of the population suffering from SAD (Furmark et al., 1999), but also a reported phenomenon in other parts of the world (Mannuzza et al., 1995). Another troubling fact is that the prevalence of anxiety is generally increasing among Swedish younger adults (Kosidou et al., 2010). This high percentage of SAD is further problematic, since symptoms often are lifelong, and that the existing treatments have a limited success rate. Speech anxiety is even more widespread than SAD, and has been identified as a problem for around one third of the total population (Ebrahimi, Pallesen, Kenter, & Nordgreen, 2019).

The most well-known and established form of psychological treatment for SAD is Cognitive Behavioral Therapy (CBT) (Olatunji, Cisler, & Deacon, 2010), a concept that is considered to be the main treatment for anxiety problems in general. A fundamental technique in CBT treatment is the patient's exposure to the feared phenomenon. This builds upon a graduated exposure to situations where the patient feels anxious, with the result of a reduced anxiety as well as reduced stress responses. Furthermore, virtual reality (VR) has been suggested as a tool for various CBT treatments to cure anxiety and other psychological disorders (Scozzari, & Gamberini, 2011).

To enhance CBT with VR has been tested for treatment of different phenomena such as agoraphobia (Castro et al., 2014), morbid obesity (Manzoni et al., 2016), and driving anxiety (Zinzow et al., 2018). There are also reports from research studies where the VR-enhanced CBT has been used to treat SAD (Ngai, Tully & Anderson, 2015), and speech anxiety (Aymerich-Franch & Bailenson, 2014; Poeschl, 2017). An advantage with VR-enhanced CBT that was pointed out by Scozzari and Gamberini (2011), is the possibility of facilitating the transition toward fearful phenomena in the real world for patients who otherwise might refuse to face the real-world stimuli.

Another emerging technology is artificial intelligence (AI), where we currently are witnessing a fast and widespread adoption of AI techniques in various daily life situations. A general critique of many AI systems is that their black box nature lacks transparency, interpretability as well as trust. A suggested way of increasing trust and to peek in to the black boxes would be to use explainable artificial intelligence (XAI). (Adadi & Berrada, 2018; Došilović, Brčić & Hlupić, 2018). In this position paper XAI is suggested to identify individuals that could profit from a VR-enhanced CTB treatment for curing speech anxiety. The aim of this paper is to present and discuss the idea of using XAI to improve CBT treatment of speech anxiety in virtual reality environments.

CBT and VR treatment of speech anxiety

The by far most frequently used psychological treatment for SAD and speech anxiety is CBT, where its fundamental principle of graduated exposure to a feared phenomenon has been proved successful in the reduction of stress and anxiety (Olatunji, Cisler & Deacon, 2010). CBT exposure treatment is based on studies of the extinction of Pavlovian fear learning. In extinction learning, a fear cue that earlier has been learned to predict an aversive event, such as an electric shock, is learned to be safe. Furthermore, it has been demonstrated that amygdala activity can be reduced during extinction learning (Ahs et al., 2015). This reduction in anxiety following exposure to public speaking was found to be associated with reduced amygdala activity (Åhs et al., 2017), indicating that the reductions in amygdala activity may be important for extinction learning and speech anxiety. A CBT reduced anxiety also reduces a patients' fear of public speaking situations and the general stress.

CBT models are based on the assumption that emotions and behaviors are influenced by how we interpret a situation. Individuals are believed to respond to cognitive representations of events and not events themselves. The consequence is that the information that reaches their consciousness does not match reality (Porto et al., 2009). CBT models for therapy are structured and targeted therapies that can vary in content. However, there are specific ingredients that most CBT therapies for SAD share. These include self-estimates, identification of negative automatic thoughts, validation, cognitive restructuring, behavioral experiments, exposure and social skills training, interspersed with clear elements of psycho-education (Rodebaugh, Holaway & Heimberg, 2004). Traditional face-to-face CBT therapies show large intergroup effect sizes, compared to waiting list, which ranges from 0.80 to 1.56 with effects persisting over long periods of time (Acarturk et al., 2014).

VR-enhanced CBT is an emerging method to recreate the anxiety felt in real-world situations to cure a patient's anxiety. A fundamental principle in exposure-based therapy is that the exposure should induce an emotional physiological reaction (Foa & Kozak, 1986). This is built on the idea that an effective VR-enhanced treatment should induce physiological reactions such as increased heart rate and skin conductance. There are also several studies indicating that therapy based on VR can produce such responses (Cote & Bouchard, 2005; Diemer et al., 2015). To summarise, VR technology seems to simulate phobic situations with enough quality to produce fear responses that corresponds to what real world situations evoke.

Finally, a VR setup would enable more exact control of the environments and also permit precise recordings of eye movements during treatment. To monitor the focus of attention of participants would add a new and interesting perspective. This is of great importance in exposure-based treatments where an obstacle to treatment success is participants that avoid looking at feared objects. Measuring eye-movements during pre-treatment could also be used to predict treatment success with the use of XAI models.

Expectancy effects on treatment

Placebo or nocebo effects are positive or negative changes in patients' symptoms attributable to the participation in therapeutic rituals, symbols, and interactions, where the treatment itself is inert (Kaptchuk & Miller, 2015). Such effects are different from those of discrete therapies, and are precipitated by the contextual or environmental cues that surround therapeutic interventions. Placebo and nocebo effects are not just attributable to normal fluctuations or regression to the mean, and they are related to biopsychosocial causes. Classical conditioning has been suggested as a model to explain this phenomenon. In the classic example, a treatment is administered by a pill where the pill will subsequently elicit a response even when the pharmacologically active ingredient is removed (Wickramasekera, 1980).

Furthermore, placebo and nocebo responses can be provoked without the participants having any prior experience. As suggested by Stewart-Williams and Podd (2004), classical conditioning can be complemented by the expectancy theory, a theory that postulates that it is the mere expectancy of an effect that produces the effect. Expectancies can be acquired in various ways, through verbal instructions, by observed experience, or by other cognitive or emotional factors (Stewart-Williams & Podd, 2004). The use of verbal instructions to influence expectancy could therefore improve the treatment effects. We suggest that AI could be used to support such verbal instructions in VR-enhanced treatments. However, how the expectancy changes when the traditional human agent providing information is replaced by an AI algorithm is not known, and is an interesting topic to investigate in the proposed research design.

Explainable Artificial Intelligence

In the current AI hype, where especially machine learning applications have superseded and exceeded human expert-judgement in many domains such as game-play (Silver et al., 2017), drug-discovery (Chen et al., 2018), and automation in fraud detection (Roy et al., 2018). One of the big downsides of those best performing methods is that they are so called 'black-box' methods that hide the complexity in multiple layers of number crunching which cannot be explained easily and may never be explainable at all (Gunning, 2017).

In areas such as public health transparency is crucial, and for support of decision making in medicine a psychologist or doctor must be able to justify a decision without just referring to a black-box model. Furthermore, a patient should have the right to know why she should carry out a risky treatment, or why she does not qualify for an expensive one. This has recently become particularly critical as a whole research track within the AI community called 'adversarial attacks' has revealed how black-box algorithms, e.g. deep learning, can be misled and completely fooled by so called adversarial attacks

(Biggio & Roli, 2018). These attacks then can be exploited in many ways, both for personal gain as well as for cyber-warfare.

Explainable AI (XAI) counteracts these downsides by 1). promoting the use of machine learning models that can be explained (Gunning, 2017; Holzinger et al., 2017), and 2). allowing the end-user of the system, in this case the psychologist, to explore properties of the model, e.g. how variables interact with one another, to understand why a decision is made. This can, for instance in diagnostics, be used to inform patients about why a treatment is expected to work for them specifically. This knowledge can be used in discussion with the patient, and thereby opens up for large possibilities in terms of transparency. For the purpose of diagnosis, a potentially large number of variables are collected from the patient. Complexity and cost of obtaining those vary largely.

Since variables can be of varying importance to the produced often non-linear diagnosis model, XAI makes it possible to analyze variables and to draw conclusions about the importance of the components, something not possible with a black-box. As a consequence, in combination with the verdict (e.g. alternatively expected treatment outcome on a scale from likely success to unlikely success), the end-user can learn about the importance of variables and possibly identify expensive variables that contribute little to the outcome.

AI derived information versus information from human agents

Placebo effects have mostly been studied in settings where the therapies are carried out by psychologists and physicians. Therefore, it has not been investigated how AI derived information could influence expectancy. However, there are a handful of studies that have investigated whether individuals process information provided by a computer differently, compared to if it is received from a human agent. In the study by (Carter et al., 2012), participants played a simplified poker game, either against computers, or against humans. During the poker games, brain functions was measured with the use of functional magnetic resonance imaging. Findings from the study showed that participants' expectancies of winning were different when they played against a computer software than what they were when they played against humans. One region of the brain that showed a strong social bias during the analysed games was the temporoparietal junction. This would indicate that this particular region in the human brain is important for changes in expectancy related to whether the opponent was a computer software or humans.

We suggest that it would be of interest to investigate if individuals that show a less strong social bias in the temporoparietal junction during this game, deem AI-derived information as more trustworthy than others. Furthermore, we also want to know whether these identified individuals benefit more from treatment by function of changes in their expectancies. Since processing of social cues robustly engages a set of brain regions, it would also be interesting to explore whether individual differences in activation of these regions predict the expectancy, and also the treatment success. This neural circuitry encompasses the lateral occipital face area, the fusiform face area, the superior temporal sulcus and the amygdala. In the recent twin study by Rosén, et al. (2020), where data were collected from around 300 twins, it was shown that individual differences in the activation of these areas also has a genetic influence. This indicates that the brain's face processing circuits may predispose some persons to be more biased towards social information than others. Therefore, some individuals could be more resistant to change expectancy from information delivered by AI algorithms.

The use of AI in treatment prediction

An emerging field in psychology is the individualisation of treatments, with the idea of improving the impact if the treatment is tailored to the individual. Here can AI play a key role in the development of new treatment strategies. Recent studies have shown that brain data collected before can be of value in the prediction of treatment success (Frick et al., 2020; Wu et al., 2020). Treatment prediction should also involve the use of AI techniques to search for patterns in large data sets. To use AI techniques, hold great promise not only in the optimisation of treatment strategies, but also for providing patients with knowledge-based predictions of treatment success. An interesting question to investigate in this proposed research is whether AI based information can be used as a treatment component to increase patients' expectation, as well as the treatment success.

The so far most stringent design for separating the true treatment-response from expectancy, is what Kirsch and Weixel (1988) named as 'balanced placebo design'. This design involves two groups that both receive the effective treatment, but only one of the groups receives correct information, while the other group is told they receive a placebo. In the suggested study we will use a variant of this design to investigate to what extent expectancy can be increased by AI prognosis. Furthermore, we will compare this expectancy boosted form of treatment to individuals that are predicted to be as likely to benefit from treatment, but are given the information that they are expected to have an average treatment effect.

These two groups should also be compared to two groups that are predicted to have average or below average treatment success. Of these two groups, one group is provided with the false information that they are very likely to benefit from treatment, while the other group is provided the information that that they have an average success-rate. Using this design, it is possible to determine the effect of correctly giving information about having high chance of treatment success.

Treatment of speech anxiety in VR environments

The suggested VR-enhanced treatment design should follow the description by Lindner et al. (2019), and be grounded in learning theory and cognitive psychology. A program with a focus on the promotion of inhibitory learning by disproving catastrophic beliefs. This should be conducted through patients' exposure to distressful speech situations in VR environments. The concept also targets self-attention, and biased recall through mental imagery-enhanced audio feedback.

Involved speech-exposure exercises should have a duration between one and three minutes each, and be conducted with no or little preparation time to maximize anticipatory anxiety. Exercises are based on unprepared speeches since excessive speech preparation is a common safety behavior, which may reduce treatment outcomes. A typical session begins with 15 minutes of psychoeducation and identification of variations from relatively easy to extremely distressful speech situations. This is followed-up by a sequence of eight exercises with speech tasks of increasing distress. Participants then mount a VR headset, and perform the exercises while audio is recorded. Immediately afterwards, the participants should remove the VR headset and evaluate the performance together with a therapist. Finally, participants should listen to the audio recording with eyes shut to use their mental imagery, to imagine seeing themselves in the third person, in the same VR environment where the speech listened to was carried out.

The used software for the stimuli presentation in the VR-head-set is a commercial application available for purchase at digital distribution platforms. Three different scenarios should be used, firstly a meeting room with a small audience seated at close distance to the speaker. Secondly, a wedding reception with a medium audience seated at a medium distance and thirdly, an auditorium setting with a large audience seated at some distance.

XAI for identifying individuals with high probability of treatment success

The earlier described concept of XAI should be used to identify individuals that are likely to benefit from the treatment. An advantage of using XAI compared to black box AI, is that the contribution of separate variables to the performance of the AI application can be evaluated (Holzinger et al., 2017). This is a clear advantage if a certain variable is expensive to assess, but does not contribute strongly to the accuracy of predicting treatment success. In a XAI setup this variable could be omitted in the further assessments.

During the last years the classical AI methods such as decision rules or decision trees, have been extended with novel XAI methods. A cloud based XAI framework that can be use to create XAI models capable of predicting treatment outcome in VR-enhanced treatment of speech anxiety is the one provided by Google (2020). This XAI service extension framework can automatically extract meta-information about results of a learning process. Furthermore, the results are visualised to facilitate a further exploration for an in-depth understanding of the ingoing models. An example, is for the understanding of how variables interact to predict treatment success. These results have a potential to identify general response patterns, which could be of value for understanding the biological and psychological basis of individual differences in inhibitory learning and the individual response to VR-enhanced speech anxiety therapy.

Conclusion

Different psychiatric disorders such as SAD and speech anxiety are a massive problem for both individuals and the contemporary society. XAI methods have the potential to be a cost-efficient alternative to improve existing CBT treatments. Results from the proposed research design could also be a relevant contribution for the understanding of how patients perceive information based on AI compared to information coming from traditional psychology. A knowledge that can be valuable for a future administration of VR- and AI-enhanced therapies. Another important contribution of the proposed research would be to investigate the effect of providing AI-based treatment prognosis on treatment success, and to isolate how such information affects expectancy. Altogether, these expected findings could be a valuable contribution to further develop AI-enhanced CBT therapies.

References

Acarturk, C., Cuijpers, P., Van Straten, A., & De Graaf, R. (2009). Psychological treatment of social anxiety disorder: a meta-analysis. *Psychological Medicine*, *39*(2), 241-254.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138-52160

Aymerich-Franch, L., & Bailenson, J. (2014). The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison. In *Proceedings of the International Society for Presence Research Annual Conference* (pp. 173-186).

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317-331.

Carter, R. M., Bowling, D. L., Reeck, C., & Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science*, *337*(6090), 109-111.

Castro, W. P., Sánchez, M. J. R., González, C. T. P., Bethencourt, J. M., de la Fuente Portero, J. A., & Marco, R. G. (2014). Cognitive-behavioral treatment and antidepressants combined with virtual reality exposure for patients with chronic agoraphobia. *International Journal of Clinical and Health Psychology*, *14*(1), 9-17.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, *23*(6), 1241-1250.

Côté, S., & Bouchard, S. (2005). Documenting the efficacy of virtual realityexposure with psychophysiological andinformation processing measures. *Applied psychophysiology and biofeedback, 30(3)*, 217-232.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology*, *6*, 26.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210-0215). IEEE.

Ebrahimi, O. V., Pallesen, S., Kenter, R. M. F., & Nordgreen, T. (2019). Psychological interventions for the fear of public speaking: a meta-analysis. *Frontiers in Psychology*, *10*, 488.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: exposure to corrective information. *Psychological bulletin*, *99*(1), 20.

Frick, A., Engman, J., Alaie, I., Björkstrand, J., Gingnell, M., Larsson, E. M., ... & Furmark, T. (2020). Neuroimaging, genetic, clinical, and demographic predictors of treatment response in patients with social anxiety disorder. *Journal of affective disorders*, *261*, 230-237.

Furmark, T., Tillfors, M., Everz, P. O., Marteinsdottir, I., Gefvert, O., & Fredrikson, M. (1999). Social phobia in the general population: prevalence and sociodemographic profile. *Social psychiatry and psychiatric epidemiology*, *34*(8), 416-424.

Google (2020). Explainable AI. (accessed 17/05/2020) from: https://cloud.google.com/explainable-ai

Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, *2*, 2.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. *arXiv preprint arXiv:1712.09923*

Lindner, P., Miloff, A., Fagernäs, S., Andersen, J., Sigeman, M., Andersson, G., ... & Carlbring, P. (2019). Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial. *Journal of anxiety disorders*, *61*, 45-54.

Mannuzza, S., Schneier, F. R., Chapman, T. F., Liebowitz, M. R., Klein, D. F., & Fyer, A. J. (1995). Generalized social phobia: Reliability and validity. Archives of general psychiatry, 52(3), 230-237.

Manzoni, G. M., Cesa, G. L., Bacchetta, M., Castelnuovo, G., Conti, S., Gaggioli, A., ... & Riva, G. (2016). Virtual reality–enhanced cognitive–behavioral therapy for morbid obesity: a randomized controlled study with 1 year follow-up. *Cyberpsychology, Behavior, and Social Networking*, *19*(2), 134-140.

Ngai, I., Tully, E. C., & Anderson, P. L. (2015). The course of the working alliance during virtual reality and exposure group therapy for social anxiety disorder. *Behavioural and cognitive psychotherapy*, *43*(2), 167-181.

Olatunji, B. O., Cisler, J. M., & Deacon, B. J. (2010). Efficacy of cognitive behavioral therapy for anxiety disorders: a review of meta-analytic findings. *Psychiatric Clinics*, *33*(3), 557-577.

Poeschl, S. (2017). Virtual reality training for public speaking—a QUEST-VR framework validation. *Frontiers in ICT*, *4*, 13.

Porto, P. R., Oliveira, L., Mari, J., Volchan, E., Figueira, I., & Ventura, P. (2009). Does cognitive behavioral therapy change the brain? A systematic review of neuroimaging in anxiety disorders. *The Journal of neuropsychiatry and clinical neurosciences*, *21*(2), 114-125.

Rodebaugh, T. L., Holaway, R. M., & Heimberg, R. G. (2004). The treatment of social anxiety disorder. *Clinical Psychology Review*, *24*(7), 883-908.

Rosén, J., Kastrati, G., Kuja-Halkola, R., Larsson, H., & Åhs, F. Genetic influence on brain responses to social and proximal stimuli. Submitted Manuscript.

Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018). Deep learning detecting fraud in credit card transactions. In *2018 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 129-134). IEEE.

Scozzari, S., & Gamberini, L. (2011). Virtual reality as a tool for cognitive behavioral therapy: a review. In *Advanced computational intelligence paradigms in healthcare 6. Virtual reality in psychotherapy, rehabilitation, and assessment* (pp. 63-108). Springer, Berlin, Heidelberg.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Chen, Y. (2017). Mastering the game of go without human knowledge. *nature*, *550*(7676), 354-359.
Stewart-Williams, S., & Podd, J. (2004). The placebo effect: dissolving the expectancy versus conditioning debate. *Psychological bulletin*, *130*(2), 324.

Wickramasekera, Ian. "A conditioned response model of the placebo effect." *Biofeedback and self-regulation* 5, no. 1 (1980): 5-18.

Wu, W., Zhang, Y., Jiang, J., Lucas, M. V., Fonzo, G. A., Rolle, C. E., ... & Wright, R. (2020). An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnology*, *38*(4), 439-447

Zinzow, H. M., Brooks, J. O., Rosopa, P. J., Jeffirs, S., Jenkins, C., Seeanner, J., ... & Hodges, L. F. (2018). Virtual reality and cognitive-behavioral therapy for driving anxiety and aggression in veterans: a pilot study. *Cognitive and behavioral practice*, *25*(2), 296-309.

Åhs, F., Kragel, P. A., Zielinski, D. J., Brady, R., & LaBar, K. S. (2015). Medial prefrontal pathways for the contextual regulation of extinguished fear in humans. *Neuroimage*, *122*, 262-271.

Åhs, F., Gingnell, M., Furmark, T., & Fredrikson, M. (2017). Within-session effect of repeated stress exposure on extinction circuitry function in social anxiety disorder. *Psychiatry Research: Neuroimaging*, *261*, 85-90.