Towards Healthcare Screening in the Wild

Stefan Scherer
University of Southern California
Playa Vista, CA 90094, USA
scherer@ict.usc.edu

Abstract

Mental health problems and problems with social functioning continue to exact a large toll on society. In the recent decades, several psychiatric diseases have accounted for an increasing percentage of global health care costs. Advances in computational behavior analyses and healthcare screening technologies provide an opportunity to create a profound impact on science and the treatment of mental health problems. Automatic algorithms to detect depression from individuals' nonverbal and verbal behavior have been lauded as a way to increase objectivity, accessibility, and acceptance. However, these advanced methods are only useful for screening for health-related problems if they can be applied in broad contexts. For accessibility to be increased, these technologies have to be deployable widely. To truly reach the masses healthcare screening tools need to work when data is collected naturalistically in the wild (i.e., online) through individuals' personal mobile devices - outside artificial and controlled conditions of the laboratory.

Author Keywords

Healthcare; Behavior Analytics

Introduction

Mental health problems continue to exact a large toll on society [1]. In the recent decades, several psychiatric diseases increased in the percentage of the global cost of disease they account for[2]. Likewise, poor social connection (i.e., loneliness) is associated with a slew of health risks from poor cardiovascular health to reduced immune system functioning [4]. Computational advances and novel ways of gathering, processing, and managing data provide an opportunity to create a profound impact on science and the treatment of mental and social health problems. More and more commonly used computational methods are beginning to be applied in mental healthcare and behavioral research [12, 5]. Such innovations are being used for screening for mental health problems, as well as other health-related issues. For example, algorithms to detect depression from audio- and video-channels have been lauded as a way to increase accessibility, overcome stigma, and ultimately helping more people get screened [12, 5, 18, 17].

However, these advanced methods - primarily investigated in controlled laboratory settings - are only useful for screening for health problems or social dysfunction if they can be applied in broader naturalistic contexts. For accessibility to be increased, for more people to get screened, these technologies have to be deployable to a number of contexts, such as a variety of healthcare facilities, offices, labs, and clinics. Beyond that, to truly reach the masses, such healthcare screening tools need to work when audio- and video-channels are collected online, through either individuals' personal computers or their mobile devices. These approaches must be capable of obtaining clear behavioral signals, in spite of background noise and other conditions, that present unique sensing and inference challenges when compared with data obtained in the controlled laboratory conditions that were used in the original development and assessment of such algorithms.

Research Agenda

For healthcare screening to be successful it needs to be widely accessible. Mobile devices and pervasive access to the Internet via smartphones rank among the most disruptive technologies of the 21st century and large parts of the population have access to it. The Pew Research Center has identified that 64% of all Americans own smartphone devices with only limited variance across a number of socio-economic variables including gender, age, income, education, as well as ethnicity and geography[3]. Through these means, we can explicitly seek to include new and diverse populations from a range of cultural, socio-economic, and geographic backgrounds as such an approach is not limited to restrictive laboratory settings. The opportunity to test automatic behavioral analyses technologies on these diverse populations is a challenging task with many complex variables. However, individuals from lower income families and rural areas are in fact the biggest potential beneficiaries of comprehensive low-cost healthcare screening, as these parts of society often suffer from significant barriers to care and help-seeking. In particular, social and cultural stigma, limited access, as well as availability of necessary

[1] Insel, T. (2011). Director's Blog: The Global Cost of Mental Illness http://www.nimh.nih.gov/about/director/2011/the-global-cost-of-mental-illness.shtml.

[2] World Economic Forum The Global Economic Burden of Non-Communicable Diseases, http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf and World Health Organisation Global Status Report on Noncommunicable Diseases, http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf.

[3] Smith, A. & Page, D. (2015). U.S. Smartphone Use in 2015. Pew Research Center Report. http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf.

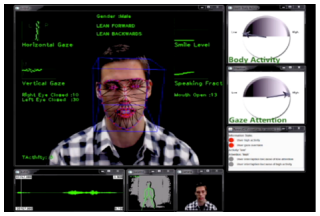

Figure 1: Visualization of MultiSense Framework.

Figure 2: SimSensei Virtual Human Interviewer.

healthcare services rank among the top. We aim to alleviate these unresolved issues through the use of novel web-based technology and an interdisciplinary approach to expand comprehensive healthcare screening at minimal cost. In the following, we introduce some of our key technologies that further this research agenda.

MultiSense

The software platform MultiSense, enables the acquisition, integration, and real-time evaluation of multimodal human behavior, which is essential for behavior analytics. The most recent version of MultiSense integrates many technologies developed by our team including the CLNF FaceTracker [2, 3] for facial landmark tracking (66 facial feature points), GAVAM HeadTracker [15] for 3D head position and orientation, head and arm gesture recognition [14] and acoustic analysis of vocalizations [10, 16]. MultiSense also integrates commercial hardware such as Microsoft Kinect sensor. For the processing of the speech signals, we use our freely available COVAREP toolbox, a collaborative speech analysis repository [7]. COVAREP provides an extensive selection of open-source robust and tested speech processing algorithms enabling comparative and cooperative research within the speech community.

SimSensei

SimSensei is a virtual human that interviews people (VH-interviewer) and identifies verbal and nonverbal indicators of mental illness. Upon completion of distress questionnaires, participants were asked to sit down in a chair facing the VH-interviewer directly, which was displayed on a large 50-inch monitor at about 1.5m distance. Within this work we utilize the SimSensei virtual human platform designed to create an engaging interaction through both verbal and nonverbal communicative channels [8]. The interaction between the participants and the fully automatic virtual human was designed as follows: the virtual human explains the purpose of the interaction and that it will ask a series of questions. It further tries to build rapport with the participant in the beginning of the interaction with a series of ice-breaker questions about Los Angeles, the location of the recordings. Then SimSensei follows with a series of more personal, open-ended clinically oriented questions with varying polarity. The agent asks both positive questions like: “What would you say are some of your best qualities?” or “What are some things that usually put you in a good mood?”, as well as more negative questions such as: “Do you have disturbing thoughts?” or “What are some things that make you really mad?”. Final “cool down” questions were also asked like: “Tell me about something you did recently that you really enjoyed?” or “When was the last time you felt really happy?”. This interview takes from 30-60 minutes, depending on the participant. However, for the purposes of the proposed work, 6 of these questions will be selected for SimSensei to ask participants online, taking about 12 minutes (2 minutes per question). The first question will serve as a proxy for the icebreaker phase, questions 2 through 5 were selected from the personal open-ended questions, and one final “cool down” question. Like interacting with a human interviewer, VH-interviewers can engage in rapport-building; and, like traditional computer-administered assessments, VH-interviewers can also increase willingness to disclose by anonymity. Specifically, our research has found that, while interacting with a VH-interviewer, holding the belief that the VH is run automatically by the computer - without oversight by a human being - will increase willingness to disclose. Specifically, interacting with VH-interviewers reduces psychological barriers to disclosure (fear of self-disclosure, impression management) and increases open, disclosure behavior in the interview [12].

Prior Investigations

Utilizing the above introduced technologies, we have pursued nonverbal indicators of depression severity in both community and clinical samples. Within the Detection and Computational Analysis of Psychological Signals (DCAPS) project, we analyzed nonverbal behavioral indicators for depression and post-traumatic stress disorder (PTSD) in civilian and veteran populations. We investigated the capabilities of automatic nonverbal behavior descriptors to identify behavioral indicators of depression and/or PTSD from video recordings of semi-structured interviews (Scherer et al., 2014; Stratou, et al., 2014). We found several statistically significant differences in the nonverbal behavior of subjects screened as having depression and/or PTSD as automatically measured with our behavior tracking algorithms. For example, an increased overall downward angle of the gaze could be identified. Further, we could identify on average significantly less intense smiles for subjects with psychological disorders as well as significantly shorter average durations of smiles.

In addition, subjects with psychological conditions exhibit on average longer self-touches and fidget on average longer with both hands (e.g. rubbing, stroking) and legs (e.g. tapping, shaking). Further, in Stratou, et al. (2014) we found reduced emotional variability in facial expressions in subjects with depression and PTSD and revealed that there are strong gender dependent effects within different behavioral indicators. We investigated acoustic behavioral indicators of PTSD and depression from the participants' speech (Scherer, et al., 2013; Scherer, et al., 2015). In particular, we focused on speaker-independent vocal tract features characterizing the speech on a breathy to tense voice quality dimension. Using this approach, we observed significant differences in the speakers' voice quality with respect to symptoms of depression and PTSD when compared to control participants. In particular, speakers with psychological disorders exhibit more tense voice qualities, confirming the previously published results (Flint, et al., 1993; Cummins, et al., 2015). Further, we identified that individuals suffering from depression and/or PTSD run their vowels together and cover a narrower range of frequencies (Scherer, et al., 2015).

Conclusions

For example, reduced frequency range in vowel production is a well documented speech characteristic of individuals suffering from psychological and neurological disorders, including but not limited to depression [6, 5, 17], cerebral palsy [11], amyotrophic lateral sclerosis [19], and Parkinson's disease [13]. However, the assessment and documentation of reduced vowel space often either rely on subjective assessments or on analysis of speech under constrained laboratory conditions (e.g. sustained vowel production, designed reading tasks), rendering analysis impractical and expensive [1]. Such limited and constrained approaches are at present the only ways to assess such acoustic characteristics, that would otherwise be inaccessible to the clinician.

Hence, we aim towards the development and deployment of automatic approaches to support clinicians and healthcare providers with much needed additional, quantified, and objective measures of nonverbal behavior to allow for a more informed and objective assessment of an individual's health status [9, 18].

Note of interest for DSLI

Stefan Scherer is a Research Assistant Professor at the University of Southern California (USC) and the USC Institute for Creative Technologies where he leads research projects funded by NSF and ARL. He received the de-

gree of Dr. rer. nat. from Ulm University with the grade summa cum laude. His research aims to automatically identify, characterize, model, and synthesize individuals' multimodal nonverbal behavior within both human-machine as well as machine-mediated human-human interaction, with particular application to healthcare and education (http://schererstefan.net).

DSLI offers the excellent opportunity to interact and exchange ideas with leading researchers in the field of mobile speech and interaction technologies. As discussed in this position paper the proliferation of cheap and scalable healthcare could have a major impact on society. As an experienced speech technologist and expert in automatic behavior analytics working in the applied field of healthcare I offer a unique perspective on a number of research topics relevant to DSLI.

REFERENCES

1. M. Alpert, E. R. Pouget, and R. R. Silva. 2001. Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders* 66, 1 (2001), 59–69.
2. T. Baltrusaitis, N. Banda, and P. Robinson. 2013. Dimensional affect recognition using continuous conditional random fields. In *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*.
3. T. Baltrusaitis, P. Robinson, and L.-P. Morency. under review. Continuous Conditional Neural Fields for Structured Regression. In *Proceedings of Computer Vision and Pattern Recognition*.
4. J. T. Cacioppo, M. E. Hughes, L. J. Waite, L. C. Hawkley, and R. A. Thisted. 2006. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. *Psychology and aging* 21, 1 (2006), 140.
5. N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. Quatieri. 2015. A Review of Depression and Suicide Risk Assessment using Speech Analysis. *Speech Communication* 17 (2015), 10–49.
6. J. K. Darby, N. Simmons, and P. A. Berger. 1984. Speech and voice parameters of depression: a pilot study. *Journal of Communication Disorders* 17, 2 (1984), 75–85.
7. G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer. 2014. COVAREP - A collaborative voice analysis repository for speech technologies. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*. 960–964.
8. D. DeVault, K. Georgilia, R. Artstein, F. Morbini, D. Traum, S. Scherer, A. Rizzo, and L.-P. Morency. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proceedings of SigDial 2013*. Association for Computational Linguistics, 193–202.
9. J. Gratch, L.-P. Morency, S. Scherer, G. Stratou, J. Boberg, S. Koenig, T. Adamson, and A. Rizzo. 2012. User-State Sensing for Virtual Health Agents and TeleHealth Applications. *Studies in health technology and informatics* 184 (2012), 151–157.
10. J. Kane, S. Scherer, M. Aylett, L.-P. Morency, and C. Gobl. 2013. Speaker and Language Independent Voice Quality Classification Applied to Unlabelled Corpora of Expressive Speech. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 7982–7986.

11. H.-M. Liu, F.-M. Tsao, and P. K. Kuhl. 2005. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America* 117, 6 (2005), 3879–3889.

12. G. Lucas, J. Gratch, A. King, and L.-P. Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

13. P. A. McRae, K. Tjaden, and B. Schoonings. 2002. Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research* 45, 1 (2002), 35–50.

14. L. Morency, A. Quattoni, and T. Darrell. 2007. Latent-Dynamic Discriminative Models for Continuous Gesture Recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*. 1–8.

15. L.-P. Morency, I. de Kok, and J. Gratch. 2008. Predicting Listener Backchannels: A Probabilistic Multimodal Approach. In *Proceedings of International Conference on Intelligent Virtual Agents 2008*.

16. S. Scherer, J. Kane, C. Gobl, and F. Schwenker. 2013. Investigating Fuzzy-Input Fuzzy-Output Support Vector Machines for Robust Voice Quality Classification. *Computer Speech and Language* 27, 1 (2013), 263–287. DOI: http://dx.doi.org/10.1016/j.csl.2012.06.001

17. S. Scherer, G. Lucas, J. Gratch, A. Rizzo, and L.-P. Morency. 2015. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Transactions on Affective Computing (in press; doi: 10.1109/TAFFC.2015.2440264)* (2015).

18. S. Scherer, G. Stratou, G. Lucas, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency. 2014. Automatic Audiovisual Behavior Descriptors for Psychological Disorder Analysis. *Image and Vision Computing Journal, Special Issue on Best of Face and Gesture 2013* 32, 10 (2014), 648–658.

19. G. S. Turner, K. Tjaden, and G. Weismer. 1995. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research* 38, 5 (1995), 1001–1013.