A Novel Framework for Distress Detection through an Automated Speech Processing System

Rajib Rana
University of Southern Queensland
Queensland, Australia
Rajib.Rana@usq.edu.au

Raj Gururajan
University of Southern Queensland
Queensland, Australia
Raj.Gururajan@usq.edu.au

Geraldine McKenzie
University of Southern Queensland
Queensland, Australia
Geraldine.McKenzie@usq.edu.au

Jeff Dunn
University of Southern Queensland
Queensland, Australia
Jeff.Dunn@usq.edu.au

Anthony Gray
University of Southern Queensland
Queensland, Australia
Anthony.Gray@usq.edu.au

Xujuan Zhou
University of Southern Queensland
Queensland, Australia
Xujuan.zhou@usq.edu.au

Prabal Datta Barua
University of Southern Queensland
Queensland, Australia
prabal.barua@usq.edu.au

Julien Epps
The University of New South Wales
Sydney, Australia
j.epps@unsw.edu.au

Gerald Michael Humphris
University of St Andrews
St Andrews, United Kingdom
gmh4@st-andrews.ac.uk

***Abstract*— Based on our ongoing work, this work in progress project aims to develop an automated system to detect distress in people to enable early referral for interventions to target anxiety and depression, to mitigate suicidal ideation and to improve adherence to treatment. The project will utilize either use existing voice data to assess people into various scales of distress, or will collect voice data as per existing standards of distress measurement, to develop basic computing algorithms required to detect various attributes associated with distress, detected through a person's voice in a telephone call to a helpline. This will be then matched with the already available psychological assessment instruments such as the Distress Thermometer for these persons. In order to trigger interventions, organizational contexts are essential as interventions rely on the type of distress. Therefore, the model will be tested on various organizational settings such as the Police, Emergency and Health along with the Distress detection instruments normally used in a psychological assessment for accuracy and validation. The outcome of the project will culminate in a fully automated integrated system, and will save significant resources to organizations. The translation of the project will be realized in step-change improvements to quality of life within the gamut of public policy.**

Keywords*— Speech recognition, distress identification, public policy

I. Introduction

Mental health issues are considered a serious concern in modern organizations as these issues can result in loss of valuable human resources and organizational productivity. Further, if not addressed properly, public policy debates dictate potential impacts on workers (e.g., discrimination), organizations (e.g., lost productivity), workplace health and compensation authorities (e.g., rising job stress-related claims), and social welfare systems (e.g., rising working age disability pensions for mental disorders) [1]. This has resulted in the rapid expansion of workplace interventions to address common mental health problems in the workplace setting, particularly as a means to prevent, detect, and effectively manage depression and anxiety [2-4]. Organizations attempt to address these issues by protecting mental health by reducing work–related risk factors, promoting mental health by developing the positive aspects of work as well as worker strengths and positive capacities, and addressing mental health problems among working people regardless of cause.

Mental health problems, both clinical (e.g., major depression, anxiety disorders) and sub-clinical (e.g., psychological distress), are very common in working populations. This *Debate* piece focuses on the workplace setting - and thus the *working* population. Given growing labour market flexibility and rising levels of unemployment and underemployment in many Organization for Economic Cooperation & Development (OECD) countries [6], addressing unemployment as well as work is now particularly important. In a recent review, the OECD estimated that similar proportions of the industrialized working-age populations are affected by clinical mental disorders: with point-prevalence estimates of 5% for severe mental disorders and another 15% for moderate mental disorders [1]. Among those affected, those with common mental disorders - depression, simple phobia, and generalized anxiety disorder - have the highest workforce participation rates [3]. In Australia, for example, the 2007 National Survey of Mental Health and Wellbeing estimated that 15% of the working population had a history of major depressive disorder (lifetime prevalence [7]); of these:

- 21% reported depressive symptoms in the past year and were in treatment
- 17% reported depressive symptoms in the past year and were not in treatment
- 11% were recovered and in treatment
- 52% were recovered and not in treatment.

In addition to clinical disorders, subclinical mental health problems and generalized distress are also prevalent in the working population [8]. In summary, mental health disorders and related problems represent a large and complex phenomenon in the workplace.

A substantial body of research has demonstrated the links between psychosocial working conditions—or job stressors—and worker health over the last three decades. Karasek and Theorell's demand-control model has been particularly influential [10]. This model hypothesized that high job strain, defined by a combination of low control over how the job is done in the face of high job demands, will be harmful to health. This was first demonstrated in relation to cardiovascular disease outcomes [10, 11]. Subsequent studies have found that job strain also predicts elevated risks of common mental disorders, even after accounting for other known risk factors [12-14]. While there is a considerable body of evidence supporting a dominant 'normal causation' model regarding the impact of working conditions on employee mental health, it should be noted that reversed causality, that is the impact of mental health on the assessment of working conditions can also occur. There is some evidence that working conditions and mental health influence each other reciprocally and longitudinally [15]. Systems thinking suggests bi-directional non-linear relationships [16] and better understanding of these processes using advanced analytic techniques (e.g., marginal structural modelling) and stronger study designs will undoubtedly be the subject of continuing research.

Numerous other job stressors, either individually or in combination, have been shown to influence mental health [14, 17, 18]. These include job insecurity, bullying or psychological harassment, and low social support at work, organizational injustice, and effort-reward imbalance [12, 14]. Unlike many historically prominent occupational exposures (e.g., asbestos), to which only a small proportion of the working population were exposed, *all working people* can be potentially exposed to job stressors. This means that even small increases in risk from such exposures can translate to substantial—and preventable—illness burdens. Given the population prevalence of a given exposure and the associated increase in risk for a specific outcome, the proportion of that outcome attributable to the exposure of interest can be estimated [19]. Based on job strain prevalence estimates of 18.6% in males and 25.5% in females and an odds ratio of 1.82 for job strain and depression [12], this method yielded estimates of job strain-attributable risk for depression in an Australian working population sample as 13% of prevalent depression among working males and 17% among working women [20]. More recently, comparable estimates were obtained from a study of the French working population for job strain-attributable risk for common mental disorders: 10.2–31.1% for men, 5.3–33.6% for women. Using a different approach, a New Zealand birth cohort study estimated that, at age 32, 45% of incident cases of depression and anxiety in previously healthy young workers were attributable to job stress [21]. While further research is needed to firmly establish the causality and magnitude of association of job strain and other stressor exposures in relation to common mental health problems (which would suggest that the attributable risks just presented are over-estimates), such single-exposure single-outcome estimates may also underestimate the proportion of mental health disorders attributable to job stressors, as a comprehensive estimate would account for all relevant job stressors and the full range of associated mental health outcomes [7]. In addition to depression, exposure to various job stressors has been associated with burnout, anxiety disorders, alcohol dependence, suicide and other mental health outcomes [14, 22]. As such, preventing or reducing exposure to job stressors and improving the psychosocial quality of work could prevent a substantial proportion of common mental health problems. Such improvements would benefit other health domains as well, as exposure to these same job stressors also predicts elevated risks for poor health behaviours as well as other high burden chronic illnesses, including cardiovascular disease [23, 24].

However, what is not occurring in the workplace is an automated way of detecting signs of mental health. In this study, we are providing a framework to detect distress in people using an automated system as an early developmental tool.

II. DISTRESS

A. Challemges in addressing Distress in an automated system

Distress is a complex phenomenon and most commonly measured using a self-reported distress thermometer [25, 26] ranging between 0 and 10, where 10 is the highest level of distress, and 0 indicates no distress. Other questionnaire-based methods to measure distress include Psychological Distress Inventory [27, 28], Hospital Anxiety and Depression Scale [29, 30], and General Health Questionnaire-12 [31, 32]. While these are all subjective measures, there is a distinct lack of an objective measure of distress. To the best of our knowledge, only one study attempted to detect distress from speech utilizing an emotional speech dataset [33], however, emotional states (happy, sad etc.) are distinct from distress. Furthermore, this study considered “acted” speech in a controlled experiment, whereas, we aim to identify distress from the spontaneous speech of patients in an uncontrolled natural environment (e.g., home), which may include ambient noise. In our proposed approach, when a cancer patient calls a cancer helpline, his/her voice will be assessed in real-time to extract relevant speech features and the features will be evaluated to identify the presence of distress. While privacy issues are a major concern, within the scope of this project, in order to preserve privacy, no actual conversation will be recorded.

To detect distress from speech, the project will address the following challenges:

1) discovering the optimal speech features most correlated to distress;
2) developing privacy-preserving techniques ensuring that the speech features or data from any intermediate stage cannot be reconstructed back to the original conversation;
3) developing algorithms to determine distress from speech features;
4) developing techniques to make the algorithms robust to ambient noise associated with the natural environment, and poor audio quality acquired through the phone microphone;
5) developing techniques to make the algorithms capable to process spontaneous speech; where fluency and volume cannot be controlled; and
6) converting the algorithms into an end-to-end prototype system which can determine the distress level in real-time.

To address the above six challenges, we propose a systems framework for distress detection to patients contacting call centres (cancer helplines) that is under-pinned by complex Deep Learning Neural Networks (DLNNs) and delivers a simplistic use case and seamless experience for the users. When a patient calls the cancer helpline or any such service, his or her voice will be simultaneously analyzed to determine distress. The patient will not be required to personally attend a clinic and he or she does not need to complete a scripted procedure or psychometric testing. Distress level will be determined solely by the spontaneous phone conversation and will then be displayed on the Operator's screen for inclusion in the assessment and referral process.

B. Speech Features Related to Distress

The basis for this research is that speech production has correlations with distress. When faced with a stressful situation, changes in physiology occur as a reflex. These physiological changes, like the increase in respiration rate, muscle tension, and a decrease in saliva production lead to changes in speech production [34]. For example, a more rapid respiration rate produces an increase in the amplitude of vocal fold vibration causing high "Intensity" in speech. An increase in laryngeal muscle tension may result in a "higher-pitched" vocal production. And a decrease in saliva production may alter the bandwidth of the "Formants" [1] (making narrower) due to the relatively drier surface of the vocal tract. In addition to utilizing the above-mentioned speech features our DLNNs learn new features/distributions (unsupervised method) from data to detect and quantify distress.

III. Distress Modelling using the Deep Learning

Deep Learning (DL) or Deep Learning Neural Networks (DLNNs) have revolutionized audio and image processing, being the technology behind the "Driverless Car". The superiority of this approach is the capacity to accurately model very complex relationships between features and labels. Therefore, the potential exists to use DLNN techniques to infer distress levels from speech features. The proposed distress detection framework is shown in Fig. 1.

1) *Training the Framework:* The framework, in particular, the DLNNs, need to be trained with historical data to detect and quantify distress in new and unseen speech. The network can be trained in both supervised and unsupervised manner. In supervised training, phone conversation data annotated for distress level is used to train the network. In the unsupervised training, the network learns to classify/identify distress speech by learning features related to distress from massive unlabeled phone conversation data. Unsupervised training occurs when there is a lack of the annotated data.

2) *Applying the Framework:* Once the network is trained it can be applied as shown in Fig 1b. When a person calls a dedicated helpline (for example emergency call 000), his or her voice will be isolated from the agent's voice. Features will be generated from the speech segments and will be given to the trained DLNNs for determining distress levels. The distress level will be calculated in real-time when the person speaks to the operator and will be shown on a dashboard at the end of the conversation.

A. Framework and its Robustness

The core building block of this study is that speech features have correlations with distress and this is supported by the comprehensive research presented in [34]. The project addresses the challenges in the following steps:

1- Dataset Creation: There is no existing speech dataset that can be readily used for our experimentation. We have approached few Queensland State Government or NGO agencies and they agreed to provide access to recorded phone conversations captured through the helpline (13 11 20) and the associated subjective rating of distress, which will be used to compile the experimental dataset.

2- Feature Selection: The project will explore speech features which are related to distress [34], such as fundamental frequency, intensity, articulation rate, vowel formants. Other features related to human affect such as, prosodic, cepstral, spectral, and glottal features, as well as features derived from the Teager energy operator (TEO) will be primarily used and then feature selection, fusion, and dimension reduction will be applied to identify the optimal feature set.

3- Privacy Preservation: Privacy preservation will be achieved in two ways: first, in the end product no raw conversation will be recorded or listened to determine distress. Second, a trade-off analysis will be conducted between reconstruction ability and performance of the feature set to select the best performing features that cannot be reconstructed.

4- Algorithm Development: Generative and non-generative deep learning neural networks including Conditional Variational Autoencoder (CVAE) networks [35], and Generative adversarial Networks (GAN) [36] will be explored to develop the distress detection algorithm. This will be iteratively developed with the feature selection steps to maximise the accuracy. Deep learning has revolutionised many fields including image processing, text processing and many more due to its power of modeling very complex relationships in data. Distress being a complex phenomenon, deep learning will be an ideal candidate to detect in from speech. While the project will focus on cutting edge deep learning algorithms, it will also consider classical methods like Support Vector Machine and will analyse the performance versus complexity trade-offs.

5- Obtain Robustness: Robustness will be achieved via selection of both features and algorithms, which are robust to noise. For example, Sparse Random Classifier (SRC) [37], which is intrinsically robust to noise, will be adapted to the deep learning models to obtain robustness. Similarly, features like cepstral time coefficients (CTC), or spectral normalization will be applied to obtain features robust to noise.

6- End-to-end Prototype System Development: The algorithms will be primarily developed off-line with the objective to maximise distress detection accuracy. Once developed the feature extraction steps and the distress detection algorithm will be optimised for running in real-time incurring reasonable computing resources.

[1] Each of the preferred resonating frequencies of the vocal tract.

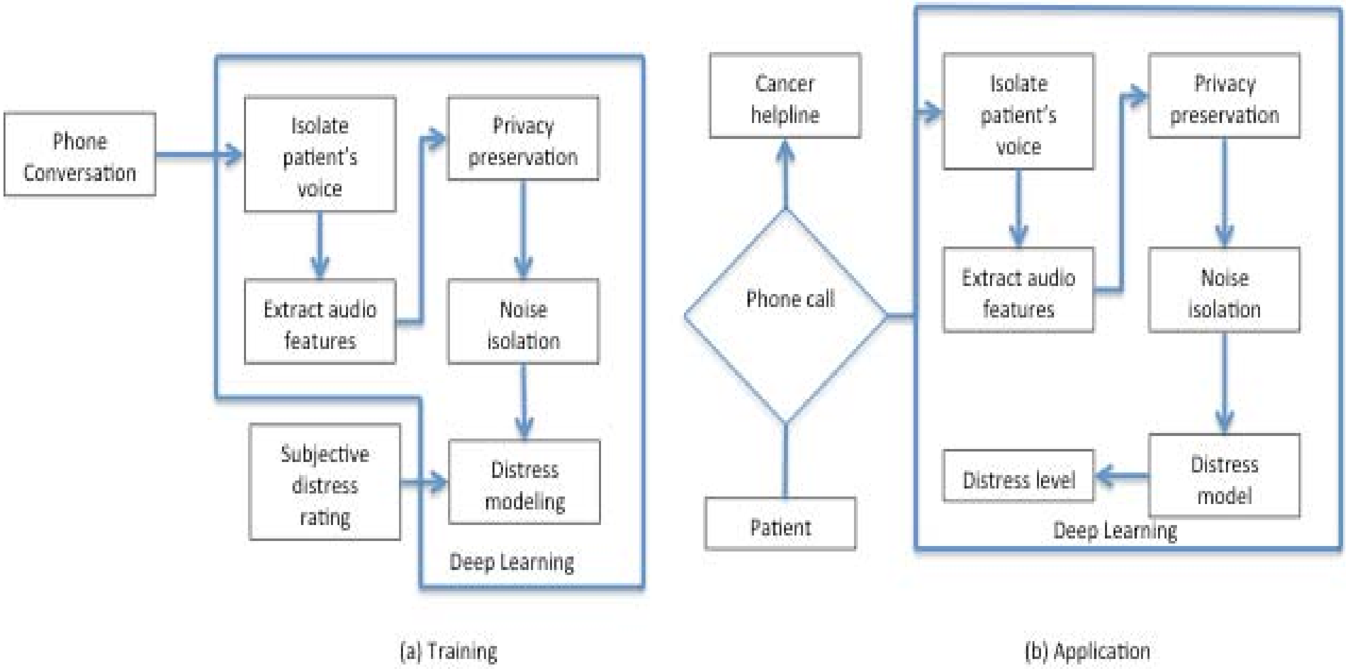

Figure 1 Framework for distress detection from phone conversation

IV. CONCLUSION AND FUTURE WORK

In this paper, we presented a framework for distress detection from phone conversation that is a part of ongoing research project entitled "an automated system to detect distress in people". The project is expected to deliver the following key outcomes:

D1 Speech features that offer the highest accuracy in distress detection subject to privacy and robustness.

D2 Robust algorithms to determine distress from speech.

D3 End-to-end prototype system to determine distress in real-time from a phone conversation.

The proposed framework achieves robustness by preserving information privacy and by being robust to ambient noise. Privacy is preserved in two ways: first, no raw conversation is recorded. Second, speech features that cannot be reconstructed are used so that even the features are compromised during processing, the original conversation can never be recovered.

Distress detection algorithms also need to be robust to ambient noise associated with the natural environment, and poor audio quality acquired through the phone microphone. Robustness is achieved via incorporation of speech features and adaptation of algorithms into the framework, which are robust to noise.

Next step, we will develop a prototype system and conduct extensive experiments based on the proposed framework. The results will be discussed and analyzed for further improving the project.

REFERENCES

[1] OECD: Sick on the job? myths and realities about mental health and work. Paris: Organisation for Economic Cooperation and Development (OECD); 2012.

[2] Wang PS, Simon GE, Avorn J, Azocar F, Ludman EJ, McCulloch J, Petukhova MZ, Kessler RC: Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes. JAMA 2007, 298(12):1401–1411.

[3] Sanderson K, Andrews G: Common mental disorders in the workforce: recent findings from descriptive and social epidemiology. Can J Psychiatry 2006, 51(2):63–75.

[4] Martin A, Sanderson K, Cocker F: Meta-analysis of the effects of health promotion intervention in the workplace on depression and anxiety symptoms. Scand J Work Environ Health 2009, 35(1):7–18.

[5] UK Government: Improving health and work: changing lives. In The Government's response to dame carol Black's review of the health of Britain's working-age population.

[6] UK: Secretaries of State of the Department for Work and Pensions and the Department of Health; 2008. Access at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210858/ hwwb-improving-health-and-work-changing-lives.pdf.

[7] OECD: OECD employment outlook 2013. In Organisation for economic cooperation & development. Paris: OECD Publishing; 2013. Access at http:// www.oecd-ilibrary.org/employment/oecd-employment-outlook-2013_ empl_outlook-2013-en.

[8] LaMontagne AD, Sanderson K, Cocker F: Estimating the Economic Benefits of Eliminating Job Strain as a Risk Factor for Depression. Melbourne: Victorian Heath Promotion Foundation (VicHealth); 2010. access at http://www. vichealth.vic.gov.au/jobstrain.

[9] Cleary C, Chant D, Wang P, Kessler R, Sheridan J, Hilton M, HA W: The prevalence of psychological distress in employees and associated occupational risk factors. JOEM 2008, 50(7):746–757.

[10] ILO: Mental Health in the Workplace. Geneva: International Labour Office; 2000. Access at http://www.ilo.org/wcmsp5/groups/public/@ed_emp/ @ifp_skills/documents/publication/wcms_108221.pdf. 10. Karasek R, Theorell T: Healthy Work: Stress, Productivity, and the Reconstruction of Working Life. New York: Basic Books, Inc., Publishers; 1990.

[11] Karasek RA, Theorell T, Schwartz JE, Schnall PL, Pieper CF, Michela JL: Job characteristics in relation to the prevalence of myocardial infarction in the US Health Examination Survey (HES) and the Health and Nutrition Examination Survey (HANES). Am J Public Health 1988, 78(8):910–918.

[12] Stansfeld SA, Candy B: Psychosocial work environment and mental health–a meta-analytic review. Scand J Work Environ Health 2006, 32(6):443–462.

[13] Bonde JP: Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. Occup Environ Med 2008, 65(7):438–445.

[14] LaMontagne AD, Keegel T, Louie AM, Ostry A: Job stress as a preventable upstream determinant of common mental disorders: a review for practitioners and policy-makers. Adv Ment Health 2010, 9(1):17–35.

[15] de Lange AH, Taris TW, Kompier MA, Houtman IL, Bongers PM: Different mechanisms to explain the reversed effects of mental health on work characteristics. Scand J Work Environ Health 2005, 31(1):3–14.

[16] Kalimo R: Reversed causality–a need to revisit systems modeling of workstress-health relationships. Scand J Work Environ Health 2005, 31(1):1–2.

[17] D'Souza RM, Strazdins L, Lim LL-Y, Broom DH, Rodgers B: Work and health in a contemporary society: demands, control, and insecurity. J Epidemiol Community Health 2003, 57:849–854.

[18] Broom DH, D'Souza RM, Strazdins L, Butterworth P, Parslow R, Rodgers B: The lesser evil: bad jobs or unemployment? A survey of mid-aged Australians. Soc Sci Med 2006, 63(3):575–586.

[19] Coughlin S, Benichou J, Weed D: Attributable risk estimation in case-control studies. Epidemiol Rev 1994, 16:51–64.

[20] LaMontagne AD, Keegel T, Vallance DA, Ostry A, Wolfe R: Job strain— attributable depression in a sample of working Australians: Assessing the contribution to health inequalities. BMC Public Health 2008, 8:9.

[21] Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE: Work stress precipitates depression and anxiety in young, working women and men. Psychol Med 2007, 37(8):1119–1129.

[22] Schneider B, Grebner K, Schnabel A, Hampel H, Georgi K, Seidler A: Impact of employment status and work-related factors on risk of completed suicide. Psychiatry Res 2011, 190(2–3):265–270.

[23] LaMontagne AD: Invited Commentary: Job strain and health behavioursdeveloping a bigger picture. Am J Epidemiol 2012, 176(12):1090–1094.

[24] Sultan-Taieb H, Lejeune C, Drummond A, Niedhammer I: Fractions of cardiovascular diseases, mental disorders, and musculoskeletal disorders attributable to job strain. Int Arch Occup Environ Health 2011, 84(8):911–925.

[25] Goldenhar LM, LaMontagne AD, Katz T, Heaney C, Landsbergis P: The intervention research process in occupational safety & health: an overview from the NORA Intervention Effectiveness Research Team. J Occup Environ Med 2001, 43(7):616–622.

[26] Jacobsen, P.B., et al., Screening for psychologic distress in ambulatory cancer patients. Cancer, 2005. 103(7): p. 1494-1502.

[27] Mitchell, A.J., Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders. Journal of Clinical Oncology, 2007. 25(29): p. 4670-4681.

[28] Morasso, G., et al., Predicting mood disorders in breast cancer patients. European Journal of Cancer, 2001. 37(2): p. 216-223.

[29] Morasso, G., et al., Assessing psychological distress in cancer patients: validation of a self-administered questionnaire. Oncology, 1996. 53(4): p. 295-302.

[30] Mitchell, A.J., N. Meader, and P. Symonds, Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. Journal of affective disorders, 2010. 126(3): p. 335-348.

[31] Zigmond, A.S. and R.P. Snaith, The hospital anxiety and depression scale. Acta psychiatrica scandinavica, 1983. 67(6): p. 361-370.

[32] Reuter, K. and M. Härter, Screening for mental disorders in cancer patients–discriminant validity of HADS and GHQ-12 assessed by standardized clinical interview. International Journal of Methods in Psychiatric Research, 2001. 10(2): p. 86-96.

[33] Goldberg, D. and P. Williams, A User's Guide to the General Health Questionnaire: GHQ (Windsor, Nelson). 350 A. Guppy and]. Marsden, 1988.

[34] Alkaher, Y., O. Dahan, and Y. Moshe. Detection of distress in speech. in Science of Electrical Engineering (ICSEE), IEEE International Conference on the. 2016. IEEE.

[35] Roberts, L.S., A forensic phonetic study of the vocal responses of individuals in distress. 2012, University of York.

[36] Kingma, D.P., et al. Semi-supervised learning with deep generative models. in Advances in Neural Information Processing Systems. 2014.

[37] Goodfellow, I., et al. Generative adversarial nets. in Advances in neural information processing systems. 2014.