ENVIRONMENTAL-RISK (E-RISK) LONGITUDINAL TWIN STUDY
CONCEPT PAPER FORM

Proposing Author(s): Dr Andre Bittar[1] and Dr Johnny Downs[1,2]

Author's affiliation, phone, and e-mail address:
[1] BRC Nucleus, King's College London
[2] Department of Child and Adolescent Psychiatry, King's College London.

Email: andre.bittar@kcl.ac.uk, johnny.downs@kcl.ac.uk

Sponsoring Investigator (if the proposing author is a student, a postdoc or a colleague): Dr Helen Fisher

Proposed co-authors: Louise Arseneault, Avshalom Caspi, Terrie Moffitt, Heidi Christensen, Nicholas Cummins, Bahman Mirheidari, Christine Aicardi

Title: Piloting development of an automated approach to coding expressed emotion in mothers' speech to improve prediction of youth mental health problems

Date: 8th June 2021

Background

For over five decades, levels of expressed emotion (EE) within families have been studied by psychologists and psychiatrists to determine which adults with mental illness are likely to have the poorest outcomes [1-3]. EE refers to the attitudes of caregivers towards their ill relative and comprises criticism, hostility, and/or emotional over-involvement as well as the degree of warmth shown towards them. EE was measured originally through in-depth face-to-face interviews [1,3] but due to time constraints has subsequently been assessed through brief samples of caregivers speaking freely about their relative (known as the Five-Minute Speech Sample [FMSS] [4]). Coding of EE from these easy-to-collect speech samples focuses on the emotions that are apparent from the way in which the caregiver speaks about their relative drawing on both what is said, and the tone of voice used. EE is one of the most robust predictors of poor outcome in the adult mental health literature, with high levels predicting greater likelihood of relapse, poorer clinical course and worse treatment response across the full spectrum of mental disorders [5,6].

Recently, the focus of EE research has shifted to prediction of mental health problems in young people [7]. Within the E-Risk Longitudinal Twin Study, the FMSS method was adapted for use with mothers of young children in the general population and it was demonstrated, using a genetically-sensitive design, that EE rated from maternal speech samples plays a causal role in the development of antisocial behavioural problems in children [8] and subsequent serious mental illnesses [9]. Other studies have shown that ratings of negative emotions from parents' speech predict the onset and course of other mental health problems in children including anxiety [10], depression [11] and attention-deficit hyperactivity disorder [12], underlining its usefulness as an early predictor of youth mental health difficulties.

However, this promising prediction method is rarely used, because the coding of speech is labour-intensive and requires highly trained raters. Moreover, human rating has limited reproducibility as it is prone to drift and unconscious biases. Recent developments in computational linguistics make it possible to represent, learn and rate natural speech using automated speech analysis [13,14]. Key advancements include bidirectional representations that analyse an utterance in the context of both preceding and following text segments [13], and the application of deep convolutional neural networks to speech recognition and synthesis [14]. The most promising approaches involve multimodal processing that integrate these methods to provide a joint representation of speech acoustics, form and content [15,16]. Multiple publications have shown that automated analysis of speech obtained from brief audio-recordings can distinguish individuals with mental disorders from unaffected controls [17,18] and accurately determine the severity of their symptoms [19,20]. For example, a recent study conducted on speech samples from 157 individuals found that multimodal representations of their speech distinguished those with depression from controls and from those with bipolar disorder with levels of balanced accuracy that were significantly above chance (0.88 and 0.75, respectively), and higher than predictions previously reported in the literature [21].

However, these methods have not been applied to predict the future development of mental disorders among children. With the additional improvements afforded by these recent developments in computational linguistics, automated speech analysis could much more efficiently estimate the likelihood that a child will develop mental health problems several years later based on an easily obtainable maternal speech sample. This has significant potential to inform the decision-making of practitioners to effectively target preventive interventions at the most vulnerable children and ultimately reduce incidence rates of mental disorders.

There is real potential for prejudices and other biases to creep into automated systems through basing the development of models on unrepresentative or skewed data. Indeed, there has been recent controversy surrounding the racially and gender-biased application of facial recognition software [22]. Thus, it is crucial for any new automated models to be developed in a responsible manner with as minimal bias as possible. In relation to speech samples, especially when coding the tone and form of words used, it will be important to test that the algorithms developed perform equally well across different socio-economic groups and dialects to reduce the likelihood of biased applications. This equality of model performance will also increase the reproducibility of the models across varying groups and contexts, thereby improving both the scientific replication of the findings and the ease of translation into research and clinical practice.

Main Objectives

This project therefore intends to achieve the following objectives:

1. Use deep learning *classification* models to detect emotional attitudes of mothers towards their children from brief samples of mothers' speech and compare with summary ratings produced by highly trained humans.
2. Test the feasibility of developing and evaluating deep learning *sequence labelling* models to annotate positive and negative comments, negativity and warmth, at a mention level which corresponds to the FMSS coding system used in E-Risk.
3. Examine the automated models' performances across socio-economic strata and geographical locations.
4. Explore whether the human ratings of expressed emotion, and deep learning classification and sequence labelling outputs are predictive of young people's mental health outcomes.

Analyses

Deep learning models

To determine whether deep learning models can detect emotional attitudes of mothers towards their children, we will use state-of-the-art methods of natural language processing (NLP) that use bidirectional representations of text and audio characteristics. These approaches have been shown to improve accuracy on numerous NLP tasks, as well as speech analysis and generation over previously established standards [13,14]. We will combine these recently developed methods in a multimodal deep learning procedure that combines information from text and audio to optimise classification and prediction [21]. In the initial model layers, we will use a pre-trained language model (e.g., BERT [13]) to represent text and will train machine learning models on extracted audio features, including the rhythm, energy, pitch and modulation of speech [14]. In a multimodal fusion layer, we will train a Long Short-Term Memory (LSTM) network to learn a joint audio-textual representation [21]. These models will be trained on two distinct tasks:

1. Classification: the first task is to directly classify speech samples across each dimension of the FMSS coding scales, thus predicting final labels for each speech sample. The bimodal fusion (audio-textual) representation will be used to predict negativism, hostility, emotional over-involvement and warmth that have already been coded by trained human raters in held-out samples of individuals 'unseen' in the model development. This is a (multilabel, multiclass) classification task.
2. Sequence labelling: the second task is to replicate the FMSS coding of speech samples as it is currently carried out by human coders. The bimodal fusion representation will be used to identify and annotate the relevant audio-textual sequences within each sample transcript in the held-out data. Final labels of the FMSS coding scales for each sample will then be derived from these identified sequences.

During model development, we will use nested cross-validation so that the majority of our sample is used for training, but the test will remain unbiased. In each iteration, approximately 80% of the total collected sample will be used for training, and 20% for testing. We will monitor prediction decrements in testing and validation samples as learning curves to diagnose underfitting and avoid overfitting in predictive models. The performance of the models will be evaluated using the area under the receiver-operating characteristic curve (AUC) to ascertain the accuracy of distinguishing speech samples with and without each specific type of emotional attitude and calibration metrics to assess agreement between observed human-rated emotional attitudes and the predicted probabilities of these derived from the automated models.

<u>Socio-economic and phenological variation in accuracy of automated output</u>

We will examine the effect of socio-economic status on the accuracy of the deep learning model outputs: by examining model performance (i.e. accuracy) stratified by major socio-economic status category groupings (at the family and neighbourhood levels). To explore geographical variation (classified by highest tier of sub-national division), and the potential contribution of accuracy impact by accent, we will examine random samples of recordings which have been stratified by accuracy performance (10 samples within 4 quintiles of low-high accuracy). Each sample will be coded by a human according to closest similarity to a reference point on a phenological variation map (23). The phenological group will be placed into 1 of 12 regional administrative groups. Fisher's exact test will be used to find differences in the distribution of regional dialects across the 4 accuracy groups.

<u>Association of human rated EE and automated outputs of the FMSS and associations with youth mental health outcomes</u>

We will explore whether the (i) human rated and (ii) automated coding of EE variables are associated with continuous measures of mental health outcomes at ages 12 and 18 years using linear regression. We will also explore how accurately the automated models of coding EE from mothers' speech can predict which children will have higher scores on dimensions of psychopathology at 12 years and at 18 years of age using the root mean square error (RMSE) and mean absolute error (MAE). Analyses will be adjusted for the non-independence of twin observations within families using the Huber-White variance estimator [24] as well as family psychiatric history and children's emotional and behavioural problems at age 5. Given the fairly modest sample size, we will focus in these pilot analyses on the continuous mental health scores rather than discrete diagnostic categories.

Variables Needed at Which Ages (names and labels):

Study: E-Risk

Age 5:

familyid	Unique family identifier
atwinid	Twin A ID (ex chkdg)
btwinid	Twin B ID (ex chkdg)
rorderp5	Random Twin Order
sampsex	Sex of Twins: In sample
sethnic	Ethnicity of twin
seswq35	Social class composite
p5cacorn	Neighbourhood deprivation at age 5
totexte5	Total Mum & Teacher Externalising Scale at 5
totinte5	Total Mum & Teacher Internalising Scale at 5

Digitised versions of Mothers' 5-minute speech samples about the twins at age 5 (currently being created by this project team)

NEGATE5	Number of negative comments - Elder twin
POSITE5	Number of positive comments - Elder twin
WARME5	Warmth towards elder twin
DISSE5	Dissatisfaction/Negativity towards elder twin
FAVORITE5	Who does mother like more?
NEGATY5	Number of negative comments - Younger twin
POSITY5	Number of positive comments - Younger twin
WARMY5	Warmth towards Younger twin
DISSY5	Dissatisfaction/Negativity towards younger twin
FAVORITY5	Who does mother like more?
TOTEE5	Total number of positive + negative EE comments - Elder - Derived by Sara Jaffee
WARMEE5	% of Positive EE Comments out of total positive + negative EE comments - Elder - Derived by Sara Jaffee
WRM50E5	Low maternal warmth - 50% or more of EE comments are -ve - Elder - Derived by Sara Jaffee

Age 12:

fhanypm12	Proportion of family members who have any mental disorder, age 12
CDIE12	Depression Scale - CDI - Elder
MASCE12	Anxiety Scale - MASC - Elder
TOTADDE12	Total Mum & Teacher ADHD Scale - Elder twin
CDTOTCRIT_EMT12	Tot CD criteria met p12_Mum or Tchr Elder, 2015
PSYSYMP01E12	Psychosis Symptom Count - Verified Coding - 0, 1+ - Elder

Age 18:

ph_e	P-factor, hierarchical, age 18
intcf_e	Internalizing, 3-factor, age 18
extcf_e	Externalizing, 3-factor, age 18
thdcf_e	Thought disorder, 3-factor, age 18

References cited:

1. Rutter M, Brown GW. The reliability and validity of measures of family life and relationships in families containing a psychiatric patient. Soc Psychiatry. 1966;1:38-53.
2. Brown GW, Birley JL, Wing JK. Influence of family life on the course of schizophrenic disorders: a replication. Br J Psychiatry. 1972;121(562):241–58.
3. Vaughn C, Leff J. The measurement of expressed emotion in the families of psychiatric patients. Br J Soc Clin Psychiatry. 1976;15:157–65.
4. Magana AB, Goldstein MJ, Karno DJ, Miklowitz JJ, Falloon IRH. A brief method for assessing expressed emotion in relatives of psychiatric patients. Psychiatry Res. 1986;17:203-12.
5. Butzlaff RL, Hooley JM. Expressed emotion and psychiatric relapse. Arch Gen Psychiatry. 1998;55:547– 52.
6. Hooley JM. Expressed emotion and relapse of psychopathology. Annu Rev Clin Psychol. 2007;3:329–52.
7. Sher-Censor E. Five Minute Speech Sample in developmental research. Dev Rev. 2015;36:127-55.
8. Caspi A, Moffitt TE, Morgan J, et al. Maternal expressed emotion predicts children's antisocial behavior problems: using monozygotic-twin differences to identify environmental effects on behavioral development. Dev Psychol. 2004;40(2):149-61.
9. Moffitt TE, Caspi A, Harrington H, Milne BJ. Males on the life-course persistent and adolescence-limited antisocial pathways: Follow-up at age 26. Dev Psychopathol. 2002;14:179–206.
10. Gar NS, Hudson JL. An examination of the interactions between mothers and children with anxiety disorders. Behav Res Ther. 2008;46,1266–74.
11. Schwartz CE, Dorer DJ, Beardslee WR, Lavori PW, Keller MB. Maternal expressed emotion and parental affective disorder: risk for childhood depressive disorder, substance abuse, or conduct disorder. J Psychiatr Res. 1990;24(3):231-50.
12. Peris TS, Hinshaw SP. Family dynamics and preadolescent girls with ADHD: The relationship between expressed emotion, ADHD symptomatology, and comorbid disruptive behavior. J Child Psychol Psychiatry. 2003;44:1177–90.
13. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018; http://arxiv org/abs/1810 04805
14. van den Oord A, Dieleman S, Zen H, et al. WaveNet: A Generative Model for Raw Audio. 2016; https://arxiv org/pdf/1609 03499
15. Li L, Zhao Y, Jiang Y, et al. Hybrid Deep Neural Network–Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition. IEEE Conference on Affective Computing & Intelligent Interaction 2013;312-7.
16. Palaskar S, Sanabria R, Metze F. End-to-End Multimodal Speech Recognition. Proc IEEE Int Conf Acoust Speech Signal Process 2018; http://arxiv.org/abs/1804.09713
17. Ooi KE, Lech M, Allen NB. Multichannel weighted speech classification system for prediction of major depression in adolescents. IEEE Trans Biomed Eng. 2013;60:497-506.
18. Stasak B, Epps J, Goecke R. An investigation of linguistic stress and articulatory vowel characteristics for automatic depression classification. Comput Speech & Lang. 2019;53:140-55.
19. Harati S, Crowell A, Mayberg H, Nemati S. Depression severity classification from speech emotion. Conf Proc IEEE Eng Med Biol Soc 2018; 2018: 5763-5766.
20. He L, Cao C. Automated depression analysis using convolutional neural networks from speech. J Biomed Inform. 2018;83:103-11.
21. Naderi H, Soleimani BH, Rempel S, Matwin S, Uher, R. Mental disorders prediction from audio speech samples using multimodal deep learning. KDD. 2019;19:doi:10.1145/1122445.1122456
22. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. Proc Mach Learn Res. 2018;81:1-15.
23. Phonological variation across the UK, British Library. https://www.bl.uk/british-accents-and-dialects/articles/phonological-variation-across-the-uk (accessed June 1, 2021).
24. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000;56(2):645-6.

Data Security Agreement

Provisional Paper Title	Piloting development of an automated approach to coding expressed emotion in mothers' speech to improve prediction of youth mental health problems
Proposing Author	Dr Andre Bittar
Today's Date	3 June 2021

Please keep one copy for your records
(Please initial your agreement)

AB__ I am familiar with the King's College London research ethics guidelines (https://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx) and the MRC good research practice guidelines (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/good-research-practice/).

AB__ My project has ethical approval from my institution.

AB__ I am familiar with the EU General Data Protection Regulation (https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-3-consent-in-research-and-confidentiality/), and will use the data in a manner compliant with its requirements.

AB__ My computer is (a) encrypted at the hard drive level, (b) password-protected, (c) configured to lock after 15 minutes of inactivity, AND (d) has an antivirus client which is updated regularly.

AB__ I will treat all data as "restricted" and store in a secure fashion.

AB__ I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.

AB__ I will not merge data from different files or sources, except where approval has been given by the PI.

AB__ I will not post data online or submit the data file to a journal for them to post.
Some journals are now requesting the data file as part of the manuscript submission process. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to the study PI for strategies for dealing with data sharing requests from Journals.

AB__ Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.

AB__ I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.

AB__ I will delete the data after the project is complete.

AB **For projects using location data:** I will ensure geographical location information, including postcodes or geographical coordinates for the E-Risk study member's homes or schools, is <u>never</u> combined or stored with any other E-Risk data (family or twin-level data)

____ **For projects using genomic data:** I will only use the SNP and/or 450K data in conjunction with the phenotypes that have been approved for use in this project at the concept paper stage.

Signature: 

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author: Dr Andre Bittar

X I have read the E-Risk data-sharing policy guidelines and agree to follow them

Provisional Paper Title: Piloting development of an automated approach to coding expressed emotion in mothers' speech to improve prediction of youth mental health problems

Potential co-authors: Johnny Downs, Helen Fisher, Louise Arseneault, Avshalom Caspi, Terrie Moffitt, Heidi Christensen, Nicholas Cummins, Bahman Mirheidari, Christine Aicardi

Potential Journals:

Intended Submission Date (month/year): April 2022

Please keep one copy for your records and return one to Louise (louise.arseneault@kcl.ac.uk)

B. To be completed by potential co-authors:

xxxx☐ Approved ☐ Not Approved ☐ Let's discuss, I have concerns

Comments: Brilliant approach to squeeze more out of the recordings on file. I am excited to see what turns up and will help as needed!

Please check your contribution(s) for authorship:

xx☐ Conceptualizing and designing the longitudinal study

xx☐ Conceptualizing and collecting one or more variables

xx☐ Data collection

☐ Conceptualizing and designing this specific paper project

☐ Statistical analyses

☐ Writing

xxx☐ Reviewing manuscript drafts

xx☐ Final approval before submission for publication

☐ Acknowledgment only, I will not be a co-author

Signature: ..Temi Moffitt...