INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

An Evaluation of the Effect of Anxiety on Speech – Computational Prediction of Anxiety from Sustained Vowels

Alice Baird[1], *Nicholas Cummins*[1,2], *Sebastian Schnieder*[3,4], *Jarek Krajewski*[4,5], *Björn W. Schuller*[1,6]

[1] Chair of Embedded Intelligence for Health Care & Wellbeing, University of Augsburg, Germany
[2] Department of Biostatistics and Health Informatics, IoPPN, King's College London, London, UK
[3] HMKW University of Applied Sciences, Berlin, Germany
[4] Institute of Experimental Psychophysiology, Düsseldorf, Germany
[5] Human-Technology Interaction, Rhenish University of Applied Science, Cologne, Germany
[6]GLAM – Group on Language, Audio, & Music, Imperial College London, UK

alice.baird@informatik.uni-augsburg.de

Abstract

The current level of global uncertainty is having an implicit effect on those with a diagnosed anxiety disorder. Anxiety can impact vocal qualities, particularly as physical symptoms of anxiety include muscle tension and shortness of breath. To this end, in this study, we explore the effect of anxiety on speech - focusing on four classes of sustained vowels (*sad*, *smiling*, *comfortable*, and *powerful*) - via feature analysis and a series of regression experiments. We extract three well-known acoustic feature sets and evaluate the efficacy of machine learning for prediction of anxiety based on the Beck Anxiety Inventory (BAI) score. Of note, utilising a support vector regressor, we find that the effects of anxiety in speech appear to be stronger at higher BAI levels. Significant differences ($p < 0.05$) between test predictions of *Low* and *High-BAI* groupings support this. Furthermore, when utilising a *High-BAI* grouping for the prediction of standardised BAI, significantly higher results are obtained for *smiling* sustained vowels, of up to 0.646 Spearman's Correlation Coefficient (ρ), and up to 0.592 ρ with all sustained vowels. A significantly stronger (Cohens d of 1.718) result than all data combined without grouping, which achieves at best 0.234 ρ.

Index Terms: anxiety disorders, sustained vowels, beck anxiety inventory, machine learning, wellbeing.

1. Introduction

Mental health can have a considerable impact on an individual's general wellbeing. In modern society, the rate of diagnosis for mental disorders characterised as anxiety disorders is increasing, particularly in urban environments [1]. Anxiety disorders refer to a subgroup of disorders which range in their severity and includes disorders such as, generalised anxiety disorder (GAD), obsessive-compulsive disorder (OCD), and post-traumatic stress disorder (PTSD). The definition of GAD (henceforth, anxiety) is, excessive worry and apprehension occurring more days than not [2]. Feelings of uncertainty often exasperate anxiety, and the current global pandemic of SARS-CoV-2 now contributes to this [3], particularly from an economic standpoint [4]. With this in mind, mechanisms to monitor and treat anxiety effectively are needed, among both general society [5], as well as for health care professionals [6].

The World Health Organisation (WHO) reports that the proportion of the global population with an anxiety disorder (as of 2015) is ca. 3.6 %, and women have a higher rate of diagnosis [7]. Known physical markers include, stomach pain and shortness of breath [8]. The Beck Anxiety Inventory (BAI) [9] is one established evaluation metric for obtaining an individual's level of anxiety. Criteria for BAI cover mental and physical characteristics and aspects which may have an effect on the vocal tract include, *difficulty in breathing* and *feeling of choking*.

Extensive and longstanding behavioural research has been made on the effect of anxiety on speech [8] and features such as, speech disturbances and varied speech-rate are amongst the characteristics of speech with which those with high anxiety tend to present [10]. Previous research suggests a redundancy in the lexical content of speech from individuals with anxiety [11]. Unlike conditions such as depression, in which research towards natural language processing approaches is becoming wide spread [12]. Additionally, acoustic aspects of speech, including disturbances and hesitations may hold meaningful information relating to anxiety [13].

In the short-term, effects of anxiety are prominent during public speaking, particularly by *social phobics*. In [14], an acoustic analysis was made of parameters including pitch, loudness, and voice quality, finding that perceived and self-assessed levels of anxiety decreased in correlation with such aspects after speaking. Similarly in [15], the authors confirm the *illusion of transparency* effect, where speakers tend to believe the prominence of anxiety in their voice is more apparent to others.

Despite much behavioural research in this area, computational approaches for monitoring and or predicting levels of anxiety are minimal. Indeed, to the best of the authors' knowledge, this study is the first to explore prediction of anxiety from adult speech. In [16], the efficacy of machine learning to monitor the speech of children with internalising disorders (including depression and anxiety) was explored. Findings show that classical acoustic approaches utilising MFCCs, and support vector machines are effective to a high degree. To this end, acoustics feature extraction toolkits including OPENSMILE and DEEP SPECTRUM have shown success for predicting similar conditions including depression [17] and stress [18].

In this study, we explore features of anxiety which may be prominent in speech and evaluate the efficacy of predicting anxiety without lexical content. We utilise various emotional classes of sustained vowels (*sad*, *smiling*, *comfortable*, and *powerful*) from the Düsseldorf Anxiety Corpus (DAC) and process the data into groupings of *Low* and *High* anxiety. As well as this, we group the aforementioned symptoms from the BAI which may explicitly effect the vocal tract – implementing both brute-force and state-of-the-art features, in a conventional support vector regressor paradigm.

http://dx.doi.org/10.21437/Interspeech.2020-1801

2. Düsseldorf Anxiety Corpus

For this study, we utilise the *Düsseldorf Anxiety Corpus (DAC)*, collected by members of the Institute of Experimental Psychophysiology, Düsseldorf, Germany. The corpus is a dataset of individuals performing various vocal exercises, featuring 252 speakers aged 18 to 68 years old (average of 31.5 years, standard deviation of 12.3 years). The files are categorised into different types of phonations, including sustained vowels, read, and free speech. The reference data is formed by measurements which includes the Beck Anxiety Inventory (BAI) [9].

We have chosen only the sustained vowels from DAC to limit the scope of the study and explore specifically the effect of anxiety without lexical content, as this has shown in previous research to be less important for anxiety [11]. For this study, we utilise four classes of sustained *a* vowels:

- **Sad** – a sad phonation of vowel [*a*] performed with low intensity and frowning face.
- **Smiling** (Smile) – a smiling phonation of vowel [*a*] in high intensity and smiling face.
- **Comfortable** (Comf) – a comfortable phonation of vowel [*a*] in comfortable intensity.
- **Powerful** (Power) – a loud phonation of vowel [*a*] in loud intensity.

These specific classes of sustained vowels are chosen due to their relation to anxiety literature – For example, negative emotion can often be masked as positive [19] (*sad*, *smiling*) and typically those with an anxiety disorder are less self-confident [20] (*comfortable*, *powerful*).

All speakers in DAC have been evaluated under the Beck Anxiety Inventory (BAI) questionnaire [21]. During the BAI, individuals answer a series of questions relating to their wellbeing, on a scale from 0–3. A total score of under 21 indicates low anxiety, and a score of above 36 indicates potentially concerning levels of anxiety. Based on these responses, we partitioned the data into two groups: *Low-BAI* vs *High-BAI* with values above 21 set as *High-BAI*. From the criteria in the BAI, a *feeling of choking* and *difficulty in breathing* are the only specific questions which may affect speech, and therefore we also group data into the presence or absence of these symptoms.

2.1. Data processing

We evaluated the dataset and selected a total of 239 speakers (69 males), which we then partitioned into train, development, and test (cf. Table 1). The audio was converted to 16 kHz, 16 bit, mono, WAV. At the beginning and end of many instances was silence. To remove this, we utilise the Librosa toolkit and automatically trim each file. The original data duration was 4 h:30 :m24 s reduced to 3 h:00 m:40 s. Within the corpus, the absolute BAI rating ranges from 0–60 and to avoid weighting for particular speakers, these raw annotations were standardised to zero mean and unit standard deviation, resulting in a range of -1.11 – +4.36.

3. Acoustic Analysis

The non-lexical acoustic information of speech is highlighted as having a stronger effect speech of high anxiety [11]. With this in mind, before proceeding with machine learning experiments, we evaluate some fundamental acoustic aspects of the samples. We extract the standard deviation (STD) of Pitch F0 (Hz), intensity (dB), and Harmonic-to-Noise-Ration (HNR) (dB) for each sample in the subset of the corpus used for our experiments. We then compare these in the *Low-BAI* and *High-BAI* pairings, through an analysis of effect size using Cohen's d, and prior to this implementing a two tailed T-test, rejecting the null hypothesis at a significance level of $p < 0.05$.

Table 1: *Speaker (#) independent partitions, Train, (Dev)elopment, and Test. Gender (M)ale:(F)emale. sustained vowel type, BAI class (Low, High), feeling of choking (No symptoms, Has symptoms), difficulty in breathing (No symptoms, Has symptoms), are reported on the audio (Inst)ance level.*

	Train	Dev	Test	$\sum$
#	74	97	68	239
M:F	26:48	25:72	19:49	69:170
Inst.	614	511	440	1565
Sustained Vowel Type				
Sad	146	127	107	380
Smile	156	124	111	391
Comf	152	130	111	393
Power	160	130	111	401
BAI Class				
Low	442	407	336	1194
High	172	104	104	371
Feeling of Choking				
No	513	456	377	1346
Has	101	55	63	219
Difficulty in Breathing				
No	460	412	351	1223
Has	154	99	89	342

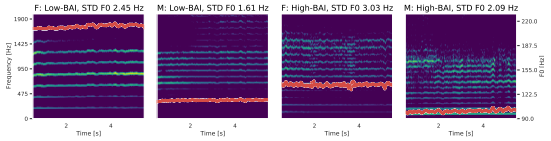
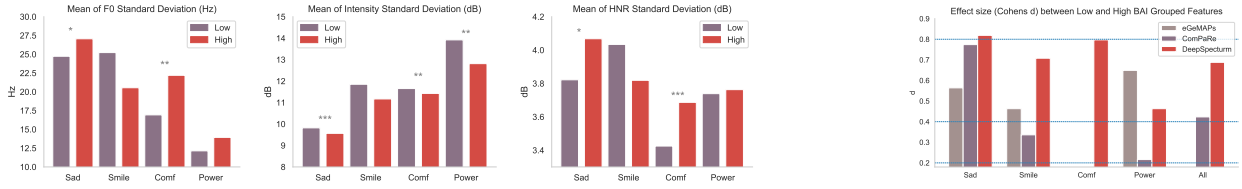


Figure 1: *Spectrogram representations of (M)ale and (F)emale samples from the* Low-BAI *and* High-BAI *grouping, vocalising the* sad *sustained vowel. F0 is plotted for each, showing a higher mean standard deviation of F0 for samples in* High-BAI *grouping as well as for Females compared to Males.*

An overview of the mean from results is given in Figure 2a. When evaluating *pitch (F0)* of the four classes of sustained vowels, we see a higher standard deviation between *Low-BAI* and *High-BAI* groupings for all classes, except *smiling* – particularly, for *sad* and comfortable, which show a smaller and medium effect size, respectively. This finding leads us to assume that lower aroused phonation types present stronger F0 variance for those with higher levels of anxiety. For the *intensity* of the speech signal, we see that in all cases samples of *Low-BAI* show strong deviation in dB, and particularly for *sad* and *powerful* which have a large and medium effect size, respectively. Additionally, we compare *Low-BAI-sad* and *Low-BAI*-power and do see a large effect size of 1.068 d, reaffirming the effect of the target vocalisation style for these sustained vowels. We also extract *HNR*, and see that like F0, all classes show higher mean results for the *High-BAI* class, aside from the *smiling* vocalisation. This finding is particularly significant for *sad* and *comfortable* and shows that vocal fold action is less consistent, for these classes in the *High-BAI* group. From a qualitative analysis of male and female speakers, we see that standard deviation in F0 appears to be larger in female speakers as shown in Figure 1, particularly for the sustained vowel class of *sad*. Due to this reason, in future studies, it would be valuable to explore genders independently.

4. Experimental Settings

We perform a series of experiments to, i) explore the efficacy of computational prediction of anxiety from speech, and ii) explore further characteristics of speech which may be affected



(a) (b)

Figure 2: *(a) shows the standard deviation for F0 (Hz), Intensity (dB) and HNR (dB) for all classes of sustained vowel, as disucssed in Section 3. In (b) the effect size between the mean of all features sets (*eGeMAPS, Deep Spectrum, *and* ComParE *for* Low-BAI *and* High-BAI *grouping of each stressed vowel type is shown, and discussed in Section 4.1. Individual results excluded reject the null-hypothesis. As marked in both (a) and (b) a small (*) effect size (Cohens d) is 0.2, medium (**) 0.4 and large above 0.8 (***).*

by anxiety. As mentioned earlier, for evaluation, we create several subsets of the data, and we evaluate each grouping through the correlation of predicted BAI score. Utilising the the four sustained vowel classes, described previously (*sad*, *smiling*, *comfortable*, and *powerful*), experiments are performed for only those speech samples, as well as for all together. Firstly we create *High-BAI* and *Low-BAI* groupings of BAI (≤ 20: Low, ≥ 21: High), we then then perform these experiments again for groups of individuals who show (Has) symptoms relating to the BAI criteria of *feeling of choking* (choking) and *difficulty breathing* (breathing) against those who do not show these systems (No).

4.1. Acoustic features

To cover a range of well-known acoustic features, we extract hand-crafted speech-based features, as well as a state-of-the-art approach, extracting spectrogram-based deep data representations from the speech signals.

openSMILE: As a conventional and well established approach, the 6 373 dimensional ComParE feature set [22], and the 88 dimensional eGeMAPS feature set [23], are used given our experience in similar paralinguistic tasks [24, 25]. From each instance, the ComParE and eGeMAPS acoustic features are extracted with the openSMILE toolkit [22]. The default parameter settings from openSMILE are used and due to the short duration of files (ca. 6 seconds), features are extracted as one feature vector per sample. We standardise the features by removing the mean and scaling to the unit variance for ComParE features – for eGeMAPS, this was not beneficial.

Deep Spectrum: Additionally, we extract a 4 096 dimensional feature set of deep data-representations using the Deep Spectrum toolkit [26][1]. Deep Spectrum has shown success for similar audio- and speech-based tasks [18], and extracts features from the audio data using pre-trained convolutional neural networks (CNNs). For this study, we extract Viridis colour map spectrograms (cf. Figure 1 for colour map), using the default VGG16 pre-trained network, and as with openSMILE, we extract one feature vector per sample. We also apply standardisation to the Deep Spectrum features.

As a brief initial step, we evaluate the effect size (Cohen's d) between *High-BAI* and *Low-BAI* groupings of the feature sets extracted for each sustained vowel (cf. Figure 2b). Of note from this analysis, we see that Deep Spectrum features appear to have consistently moderate effect sizes, larger than ComParE and eGeMAPS, particularly for the *sad* and *comfortable* class. As this finding also seems to be reflective of our previous acoustic analysis, where *sad* and *comfortable* seem to behave similarly for F0 STD and HNR STD, this leads us to assume given the visual nature of Deep Spectrum features that increased standard deviation in F0 for those with higher anxiety may be more easily captured with these features. Further to this, Deep Spectrum most likely observes noise in the signal, as reflected by the high HNR for both *sad* and *comfortable*.

[1] https://github.com/DeepSpectrum/DeepSpectrum

4.2. Training and evaluation

Given that our dataset is reasonably small (ca 3hrs), for a robust and easily reproducible approach, we choose to utilise an epsilon-support vector regressor (SVR) with a linear kernel. We split the data for training, into speaker-independent sets: training, development and test (cf. Table 1). During the development phase, we trained a series of SVR models, optimising the complexity parameters ($C \in 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1$), and evaluating their performance on the development set. We then re-trained the model with the concatenated train and development set, and evaluate the performance on the test set. We repeat this method for each combination. Note that we report the best value for C in development for test validation.

To evaluate the results of all experiments, we utilise *Spearman's correlation coefficient* (ρ) due to the ordinal nature of the raw BAI values. Additionally, we Cohen's d is used as a measure of effect size between the predictions of results of interest. Reporting of Cohen's d proceeds an evaluation of each test set prediction result for normality using a Shapiro-Wilktest [27], as well as two-tailed T-test, rejecting the null hypothesis at a significance level of $p < 0.05$.

5. Results and Discussion

Our fully-fledged results are given in Table 2. As indicated by *, there are significant difference in almost all predictions for *Low-BAI* vs *High-BAI* groupings. As well as this in most cases, *High-BAI* grouped results are significantly higher than *Low-BAI* grouped results. Although our results do vary, they suggests that the characteristics of speech, harnesses for prediction of anxiety, are stronger when anxiety is at high levels. This finding is supported by earlier discussed literature, which suggests that speech disturbances and varied speech-rate are prominent in the speech of those with high anxiety [10]

Looking closer at our BAI grouped experiments, we see *High-BAI* grouped anxiety predictions are stronger, with at best, 0.505 ρ for prediction of standardised BAI of all *High-BAI* grouped samples. Through the late-fusion of the two best results eGeMAPS and Deep Spectrum, this is increased to 0.592 ρ. For the individual sustained vowels, *smiling* in *High-BAI* grouping performs best, with eGeMAPS showing up to 0.593 ρ, a result which is also improved by late-fusion up to 0.646 ρ. We see a slight moderate correlation for Deep Spectrum of *sad High-BAI* grouping. However, this is not consistent with all feature sets. For *comfortable* and *powerful*, there are no substantial correlations, leading us to consider that these samples do not provide meaningful information for the current task.

For the grouping of *Has-symptoms*, or *No-symptoms* of

Table 2: *(Dev)elopment and Test SVR results for the prediction of standardised BAI for all stressed vowel combinations selected from DAC; Reporting Spearmans Correlation Coefficient (ρ) for groupings of;* Low-BAI *or* High-BAI*, (Has) Symptoms or (No) Symptoms of Feeling of (Choke)ing, (Has) Symptoms or (No) Symptoms of Difficulty in (Breath)ing. For* High-BAI *and Has symptoms groupings* $*$ *indicates significance ($p < 0.05$) of test predictions as compared to the equivalent* Low-BAI *or No symptoms grouping test prediction. For BAI grouped results, we include late-fusion results taken from the mean of predictions of the two best performing feature sets. Emphasised results show a positive ρ correlation above 0.3.*

ρ	**Sad**				**Smiling**				**Comfortable**				**Powerful**				**All**			
BAI	Low		High		Low		High		Low		High		Low		High		Low		High	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test
eGeMAPS	-.049	.106	.210	-.057*	.043	-.012	.181	**.593***	-.053	.029	.350	.031*	-.098	-.294	.084	.087*	.012	-.033	.173	**.405***
ComParE	.025	-.018	.100	-.241*	.015	-.271	.441	**.446***	-.101	.183	.154	-.127*	.096	.073	.252	.143*	.077	.002	.120	.120*
Deep Spectrum	.104	.210	.005	**.304***	.190	.012	.253	**.418***	.219	.216	.146	-.145*	.132	-.004	.146	-.501*	.141	.286	.105	**.506***
Late-fusion	–	.194	–	**.228***	–	.008	–	**.646***	–	.213	–	.027*	–	-.029	–	.167	–	.238	–	**.592***
Choke	No		Has		No		Has		No		Has		No		Has		No		Has	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test
eGeMAPS	.102	.190	.611	-.064*	.252	**.350**	.343	.218*	.192	.067	.132	-.424	.039	**.376**	-.148	**.317***	.146	.222	.103	-.029*
ComParE	.099	-.170	.110	.051*	.309	.011	.223	**.535***	-.008	.294	.081	-.463	.057	.201	.392	**.494***	.102	-.163	-.148	-.392*
Deep Spectrum	.078	.288	.369	-.397*	.202	.246	.022	-.160*	.130	.300	-.003	.297*	.106	**.438**	.706	.200*	.188	.254	.075	.118*
Breath	No		Has		No		Has		No		Has		No		Has		No		Has	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test	Dev	Test
eGeMAPS	-.019	.120	.276	-.122*	.187	.255	.453	**.340**	.036	.067	.413	-.217	.081	.234	.218	.194*	-.019	.120	.187	.255*
ComParE	.048	-.139	-.028	**.363***	.282	-.021	.357	**.699***	.090	.284	.082	.078*	.099	.212	.259	-.384*	.112	.136	.237	**.379***
Deep Spectrum	-.006	**.357**	.301	.284*	.342	.256	-.045	.028*	.004	**.302**	-.010	.169*	.176	**.384**	-.009	-.026*	.151	.285	.178	.126*

feeling of choking, *smiling* samples again performs best, with ComParE at best 0.535 ρ. However, in this case, eGeMAPS and Deep Spectrum are less able to capture the phenomena. *Comfortable* phonations show a strong negative correlation for the *Has-symptom* grouping, a finding which to a degree also appears for *sad*, suggesting that intensity may play a strong roll in this task. When predicting standardised BAI from all samples with *No-symptom* of *choking*, we see that this is stronger than the *Has-symptoms* pairing. Overall, there are no strong findings from this paradigm. However, most *No-symptoms* grouped results perform better than *Has-symptoms* grouped, which suggest a need for further acoustic analysis, to observe any variation in the samples for this constellation.

For the grouping of *Has-symptoms* or *No-symptoms* of *difficulty in breathing*, we see that as with *choking*, the *No-symptoms* grouped results are often stronger than *Has-symptoms* group. However, across features sets, this is somewhat confused. For *sad*, for example, the *No-symptoms* grouping performs better with Deep Spectrum , but overall, ComParE shows slightly better results for the *Has-symptoms* grouping. Like all other groupings, the *smiling* class in the *Has-symptoms* grouping shows our best result with up to 0.699 ρ. ComParE also performs best when utilising all data for the *Has-symptoms* grouping. This is suggesting that HNR, which may be stronger due to restricted airflow, is more easily captured by ComParE features for individuals with this *breath* symptom.

To evaluate further the degree to which highly anxious speech improves prediction accuracy, we additionally reran our experiments with all data and without any groupings (cf. Table 3). From this we find that still the *High-BAI* grouped with Deep Spectrum and eGeMAPS results are stronger, with the best results from late-fusion being .243 ρ for all data (a result which can be considered negligible). This result is significantly lower than the best High-BAI result with all sustained vowels, reporting a very large effect size of 1.718 d.

In general, for all scenarios, the *smiling* class performs best for the stronger *High-BAI*/*Has-symptoms* groupings. This finding could suggest that anxiety is more prevalent in a more facially strained stressed vowel. There is much in the literature relating to smiling and anxiety, for example, the "fooled by a smile" effect in which those who suffer from anxiety can show untrue emotional expressions [28]. Furthermore, high anxiety involves much more facial expression, and general movement, as compared to lower anxiety, with 'non-enjoyment' smiles being displayed frequently [29].

Table 3: *SVR results for prediction of standardised BAI from all data combined, without groupings. Late-fusion of two best.*

	Dev	Test
eGeMAPS	.189	.245
ComParE	.093	.213
Deep Spectrum	.106	.238
Late-fusion	–	**.243**

6. Conclusion and Future Outlook

In this study, we explored the effect of anxiety on speech. In particular, we evaluated the efficacy of predicting anxiety from adult speech for the first time and evaluated non-lexical sustained vowels as a first step. Our findings show that utilising speech-based features for prediction of anxiety is valid and that recognition of higher levels of anxiety is better. As individuals reporting high levels of BAI may need a more timely medical intervention this finding is promising. From our results, we see that *smiling* phonations are particularly informative for those with high anxiety. A finding related to literature which states that smiling causes an alteration of the vocal tract and can be "heard as well as seen" [30]. Additionally, those with high anxiety often overstate their emotional expression [28], possibly leading to stronger speech variance. For further studies, we hope to explore the effect of *smiling* phonations (and facial movement) on anxious speech further. As well as this, given the slight gender bias to our data, it would be of interest to evaluate gender independently, as we have seen that features such as STD for F0 are stronger in highly anxious female samples. Further, since our study shows promise for the presence of anxiety in speech without lexical content, it would be of interest to compare this to free speech samples.

7. Acknowledgements

This work is funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B), and the Horizon 2020, EFPIA funded Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 115902.

8. References

[1] K. McKenzie, A. Murray, and T. Booth, "Do urban environments increase the risk of anxiety, depression and psychosis? an epidemiological study," *Journal of Affective Disorders*, vol. 150, no. 3, pp. 1019–1024, 2013.

[2] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington DC, USA: APA, 2013.

[3] E. A. Holmes, R. C. O'Connor, V. H. Perry, I. Tracey, S. Wessely, L. Arseneault, C. Ballard, H. Christensen, R. C. Silver, I. Everall *et al.*, "Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science," *The Lancet Psychiatry*, 2020, in press, 14 pages.

[4] S. R. Baker, N. Bloom, S. J. Davis, and S. J. Terry, "COVID-Induced Economic Uncertainty," National Bureau of Economic Research, 2020, 17 pages. [Online]. Available: https://www.nber.org/papers/w26983

[5] M. Shevlin, O. McBride, J. Murphy, J. G. Miller, T. K. Hartman, L. Levita, L. Mason, A. P. Martinez, R. McKay, T. V. Stocks *et al.*, "Anxiety, Depression, Traumatic Stress, and COVID-19 Related Anxiety in the UK General Population During the COVID-19 Pandemic," PsyArXiv, 2020, 27 pages. [Online]. Available: https://psyarxiv.com/hb6nq

[6] T. Shanafelt, J. Ripp, and M. Trockel, "Understanding and Addressing Sources of Anxiety Among Health Care Professionals During the COVID-19 Pandemic," *JAMA*, 2020, 2 pages. [Online]. Available: https://doi.org/10.1001/jama.2020.5893

[7] World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," WHO, 2017, 24 pages. [Online]. Available: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

[8] B. Pope, T. Blass, A. W. Siegman, and J. Raher, "Anxiety and depression in speech," *Journal of Consulting and Clinical Psychology*, vol. 35, no. 1p1, p. 128, 1970.

[9] T. Fydrich, D. Dowdall, and D. L. Chambless, "Reliability and validity of the beck anxiety inventory," *Journal of Anxiety Disorders*, vol. 6, no. 1, pp. 55–61, 1992.

[10] M. Cook, "Anxiety, speech disturbances and speech rate," *British Journal of Social and Clinical Psychology*, vol. 8, no. 1, pp. 13–21, 1969.

[11] L. A. Gottschalk and E. C. Frank, "Estimating the magnitude of anxiety from speech," *Behavioral Science*, vol. 12, no. 4, pp. 289–295, 1967.

[12] Z. Huang, J. Epps, D. Joachim, and V. Sethu, "Natural language processing methods for acoustic and landmark event-based features in speech-based depression detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 435–448, 2020.

[13] S. V. Kasl and G. F. Mahl, "Relationship of disturbances and hesitations in spontaneous speech to anxiety," *Journal of Personality and Social Psychology*, vol. 1, no. 5, pp. 425—-433, 1965.

[14] P. Laukka, C. Linnman, F. Åhs, A. Pissiota, Ö. Frans, V. Faria, Å. Michelgård, L. Appel, M. Fredrikson, and T. Furmark, "In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech," *Journal of Nonverbal Behavior*, vol. 32, no. 4, pp. 195–214, 2008.

[15] A. M. Goberman, S. Hughes, and T. Haydock, "Acoustic characteristics of public speaking: Anxiety and practice effects," *Speech communication*, vol. 53, no. 6, pp. 867–876, 2011.

[16] E. W. McGinnis, S. P. Anderau, J. Hruschak, R. D. Gurchiek, N. L. Lopez-Duran, K. Fitzgerald, K. L. Rosenblum, M. Muzik, and R. S. McGinnis, "Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2294–2301, 2019.

[17] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[18] A. Baird, S. Amiriparian, N. Cummins, S. Sturmbauer, J. Janson, E.-M. Messner, H. Baumeister, N. Rohleder, and B. Schuller, "Using Speech to Predict Sequentially Measured Cortisol Levels During a Trier Social Stress Test," in *Proc. INTERSPEECH 2019*. Graz, Austria: ISCA, 2019, pp. 534–538.

[19] C. Dijk, A. H. Fischer, N. Morina, C. van Eeuwijk, and G. A. van Kleef, "Effects of Social Anxiety on Emotional Mimicry and Contagion: Feeling Negative, but Smiling Politely," *Journal of Nonverbal Behavior*, vol. 42, no. 1, pp. 81–99, 2018.

[20] S. Hanton, S. D. Mellalieu, and R. Hall, "Self-confidence and anxiety interpretation: A qualitative investigation," *Psychology of Sport and Exercise*, vol. 5, no. 4, pp. 477–495, 2004.

[21] A. T. Beck, N. Epstein, G. Brown, and R. A. Steer, "An inventory for measuring clinical anxiety: Psychometric properties," *Journal of Consulting and Clinical Psychology*, vol. 56, no. 6, pp. 893—-897, 1988.

[22] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor," in *Proc. ACM MM '13*. Barcelona, Spain: ACM, 2013, pp. 835–838.

[23] F. Eyben, K. Scherer, B. Schuller, J. Sundberg, E. André, C. Busso, L. Devillers, J. Epps, P. Laukka, S. Narayanan, and K. Truong, "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[24] A. Baird, S. Amiriparian, N. Cummins, A. M. Alcorn, A. Batliner, S. Pugachevskiy, M. Freitag, M. Gerczuk, and B. Schuller, "Automatic Classification of Autistic Child Vocalisations: A Novel Database and Results," in *Proc. INTERSPEECH 2017*. Stokholm, Sweden: ISCA, 2017, pp. 849–853.

[25] A. Baird, S. Amiriparian, and B. Schuller, "Can Deep Generative Audio be Emotional? Towards an Approach for Personalised Emotional Audio Generation," in *Proc. MMSP*. Kuala Lumpur, Malaysia: IEEE, 2019, 5 pages.

[26] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, and B. Schuller, "Snore Sound Classification Using Image-based Deep Spectrum Features," in *Proc. INTERSPEECH 2017*. Stockholm, Sweden: ISCA, 2017, pp. 3512–3516.

[27] J. Peat and B. Barton, *Medical Statistics: A Guide to Data Analysis and Critical Appraisal*. Malden, MA, USA: Blackwell Publishing Ltd, 2008.

[28] J. A. Harrigan and K. T. Taing, "Fooled by a Smile: Detecting Anxiety in Others," *Journal of Nonverbal Behavior*, vol. 21, no. 3, pp. 203–221, 1997.

[29] J. A. Harrigan and D. M. O'Connell, "How do you look when feeling anxious? Facial displays of anxiety," *Personality and Individual Differences*, vol. 21, no. 2, pp. 205–212, 1996.

[30] V. C. Tartter, "Happy talk: Perceptual and acoustic effects of smiling on speech," *Perception & Psychophysics*, vol. 27, no. 1, pp. 24–27, 1980.