Spotting the Traces of Depression in Read Speech: An Approach Based on Computational Paralinguistics and Social Signal Processing

Fuxiang Tao[1]*, Anna Esposito*[2]*, Alessandro Vinciarelli*[1]

[1]University of Glasgow, Glasgow, UK
[2]Universitá della Campania "Luigi Vanvitelli", Caserta, Italy

f.tao.1@research.gla.ac.uk, Anna.Esposito@unicampania.it,
Alessandro.Vinciarelli@glasgow.ac.uk

Abstract

This work investigates the use of a classification approach as a means to identify effective depression markers in read speech, i.e., observable and measurable traces of the pathology in the way people read a predefined text. This is important because the diagnosis of depression is still a challenging problem and reliable markers can, at least to a partial extent, contribute to address it. The experiments have involved 110 individuals and revolve around the tendency of depressed people to read slower and display silences that are both longer and more frequent. The results show that features expected to capture such differences reduce the error rate of a baseline classifier by more than 50% (from 31.8% to 15.5%). This is of particular interest when considering that the new features are less than 10% of the original set (3 out of 32). Furthermore, the results appear to be in line with the findings of neuroscience about brain-level differences between depressed and non-depressed individuals.

Index Terms: Computational Paralinguistics, Social Signal Processing, Depression Detection, Read Speech

1. Introduction

Extensive surveys of the population suggest that 16.2% of the adults in the USA experience at least one episode of Major Depression Disorder (MDD) during their life [1]. Moreover, the pathology appears to be associated with more than half of all suicides [2] and it is one of the main causes of disability among people above 5 years of age [3]. As a result, depression is a major economic and societal burden. For example, in Canada, the average medical costs for depressed people are 3.5 times higher than those for non-depressed ones [4]. Similarly, in the USA, the costs associated to depression (medical expenses, productivity loss, etc.) sum up to USD 70 billions per year [5].

However impressive, the figures above are still likely to be an underestimate because only half of the patients obtain medical attention and, furthermore, only 21% of them undergo adequate treatment [1]. In other words, while being a serious pathology with highly negative consequences, depression tends to remain undetected or to be poorly treated. One possible reason of such a situation is that the first line of intervention against depression is not led by psychiatrists, specialised and experienced in the diagnosis of mental health issues, but by *General Practitioners* (GP), doctors expected to deal with common pathologies, but to refer to specialists in case of more serious problems. In particular, due to the difficulties in diagnosing depression, the accuracy of GPs has been shown to range between 57.9% and 73.1% for the data used in the experiments of this work, thus leaving a large number of cases undetected [6].

One possible way to address the problems above, at least to a partial extent, is to identify *depression markers*, i.e, reliable, measurable and, possibly, machine detectable traces of the pathology. In fact, the availability of such markers can make it easier for non specialised doctors to effectively identify people affected by depression and, correspondingly, to increase the percentage of cases that obtain psychiatric attention and proper treatment. For these reasons, this article investigates the use of computational paralinguistics [7] and social signal processing [8] as a means to identify depression markers in the way people read a predefined text.

The main reason for focusing on read speech is that asking potential depression patients to read a text is something that can be done easily in a clinical setting. This is a major advantage with respect to biological markers investigated so far that require invasive exams like, e.g., brain neurotrophic factors [9] or monoamine levels in cerebrospinal fluids [10]. Not to mention that these markers have been investigated while developing pharmacological treatments and have been shown to be affected by several drawbacks, including the difficulty to interpret placebo-controlled trials [11, 12] methodological pitfalls at the level of patient selection and enrolment [13] or misalignment between clinicians' observations and patients' self-assessments [14]. Similarly, the efforts of the computing community have explored a wide spectrum of behavioural markers (facial expressions, nonverbal vocal behaviour, etc.), but none of them appear to clearly outperform the others. Furthermore, compared to the analysis of read speech, other behavioural cues might be difficult to capture and analyse outside a laboratory setting.

The experiments of this work have involved 110 participants that have been recorded while reading the same text. An approach based on a standard feature set, originally designed to recognise emotions [15], has been used to perform initial experiments. The feature set has then been expanded with a few features accounting for two main behaviours, namely reading speed and use of silences. The main reason for focusing on such behaviours is that, according to neuroscience, one of the main effects of depression is that the brain tends to become slower at processing linguistic information. Therefore, it is reasonable to expect that depressed individuals tend to read slower and to be less fluent. The difference in performance resulting from the expansion of the feature set has been used as a measure of how effectively the behaviours above can account for the presence of depression. The results show that adding 3 features to the initial 32 is sufficient to increase the accuracy from 68.2% to 84.5%, corresponding to a reduction of the error rate by 51.2%. In other words, reading speed and silences appear to be reliable markers of depression.

The rest of this article is organised as follows: Section 2 surveys previous work, Section 3 describes the data used in the experiments, Section 4 reports on experiments and results, while

Sections 5 draws some conclusions.

2. Previous Work

The computing community has made substantial efforts towards the automatic detection of depression in speech (see [16] for an extensive survey). Overall, two main tasks have been addressed. The first is the inference of scores obtained through the administration of self-assessment questionnaires (e.g., the *Beck Depression Inventory II*). The second is actual depression detection, i.e., automatic discrimination between people diagnosed with depression by psychiatrists and control individuals that are not affected by mental health issues. In a few cases, the efforts have targeted the identification of depression markers like in this work.

In several cases, the proposed approaches do not model only the speech signals but also their transcriptions [17, 18, 19, 20]. In particular, the suggestion proposed in [17] is that acoustic and linguistic aspects of speech should never be addressed separately in depression related technologies. However, the results of other works show that the best performances result from the use of transcriptions only [18]. Furthermore, according to the experiments in [19], the multimodal combination of paralinguistics and text can work only when using deep networks with attention gates [19]. Finally, in the case of the experiments in [20], the indication is that the best results can be obtained only when taking into account interaction dynamics, i.e., when a given sentence is uttered in a conversation.

Overall, the results above suggest that it is unclear whether taking into account what people say actually helps or not. However, the experiments of this work are based on read speech and all participants utter the same words in the same order. Therefore, linguistic aspects of the data cannot contribute to the discrimination between depressed and non depressed participants. As a consequence, the focus is on the sole speech signal like in a large number of other contibutions including, e.g., [21, 22, 23]. In the first work [21], the experiments aim at testing whether speech samples captured through mobile phones allow one to discriminate between people that are above or below a threshold score of the *Personal Health Questionnaire*. The results show that this is actually possible with an accuracy of 72%.

In the other two works [22, 23], the goal is to identify depression markers that, like in this work, can help to distinguish between depressed individuals and the others. The focus of the experiments in [22] is on adolescents because their voice is not fully formed and, therefore, speech-based depression detection can be a more challenging task. The results of the work show that the most effective marker is the energy of the signal, corresponding to how loud people speak, especially when measured with the Teager operator [7]. In contrast, the marker that appears to be more effective in [23] is the variability of phonetic characteristics.

3. The Data

The experiments of this work have involved 110 people, including 54 individuals that have never experienced mental health issues, referred to as *control participants*, and 56 depression patients. The control participants were recruited via a word of mouth process, while the depressed ones were recruited among the patients treated in three mental health centres in Southern Italy. All depressed participants have been diagnosed by professional clinicians with one of the following pathologies: Major Depressive Disorder (19 cases), bipolar disorder in depressive phase or with last depressive episode (13 cases), reactive depression (7 cases), endo-reactive depression (6 cases) and anxiety-depressive disorder (4 cases). No specific diagnosis was provided for the remaining 7 patients. All participants are native Italian speakers and have been asked to read aloud a tale by Aesop (The North Wind and the Sun).

Table 1: *Participant demographics. In the table,* M *stands for Male,* F *stands for Female,* L *for Lower Education and* H *for Higher Education. The total across the education levels is 105 because 5 participants did not provide information about their studies.*

Condition	Age	M	F	L	H
Control	47.6 ± 12.6	12	42	19	33
Depressed	47.2 ± 12.3	18	38	27	26
Total	51.2 ± 15.9	30	80	46	59

Table 1 provides demographic information and shows there is no difference between control and depressed participants in terms of age distribution ($p > 0.05$ according to a two-tailed t-test), gender balance ($p > 0.05$ according to a χ^2 test) and distribution across the two main education levels in Italy ($p > 0.05$ according to a χ^2 test), namely Lower (at most 8 years of study) and Higher (at least 13 years of study). The above is important because it shows that any detectable differences between the two groups of participants are likely to result from their condition (depressed or control) and not from other factors that might influence the way people read.

The reason why the number of female participants is significantly higher is that women tend to develop depression more frequently than men [24]. Therefore, the gender distribution of the corpus is more representative of what is observed among depression patients. Similarly, the age range is the same as the one observed across depression patients. In this respect, the corpus is designed to be as representative as possible of the typical patients of the pathology.

4. Experiments and Results

The baseline approach used in the experiments follows the methodologies of computational paralinguistics [7]. In particular, a speech signal is segmented into 25 *ms* long analysis windows (or *frames*) that start at regular time steps of 10 *ms*. Each frame is mapped into a 32-dimensional feature vector (see below for more details) and the speech signal is then represented with the average of the vectors extracted from the individual frames. Such an average is then fed to a classifier that assigns the speech signal to one of the two possible classes, namely *depressed* or *control*. The features correspond to those designed for the *Interspeech 2009 Emotion Challenge* [15]. The main motivation is that such a set has been successful in a wide spectrum of problems - especially when it comes to the inference of psychological information from speech and, therefore, can be considered as a standard baseline for comparison. The feature set builds upon 16 core features:

- Root Mean Square of the Energy (Energy): it accounts for how loud someone speaks and it is known to have an association with depression (see Section 2);
- Mel-Frequency cepstral coefficients 1-12 (MFCC): they account for the phonetic content of the data;
- Fundamental Frequency (F0): is the frequency that car-

Table 2: *Results of baseline classifier.*

Classifier	Features	Accuracy	Precision	Recall
Baseline	32	68.2%	68.2%	68.2%

Table 3: *Performance after taking reading speed into account.*

Classifier	Features	Accuracy	Precision	Recall
Speed	33	77.3%	77.3%	77.2%

ries the highest energy in the signal, known to shift to higher bands in case of depressed speakers [25, 26, 27];

- Zero-Crossing Rate (ZCR): it accounts for F0 and it is reported to lead to 80% accuracy in predicting whether listeners without medical background consider a speaker depressed [28];
- Voicing probability (VP): it accounts for the probability of a frame corresponding to emission of voice, it has been shown to capture information about several affective states [29, 16].

Since the features above are extracted from every frame, the set can be enriched by adding the difference between the value of every feature in the current frame and the value in the previous frame. In this way it possible to take into account how the features change over time and the final feature set includes 32 elements. The feature extraction has been performed with *OpenSMILE* [30].

The feature vectors extracted from the 110 recordings have been fed to a linear kernel SVM trained according to a leave-one-out experimental design, meaning that an SVM has been trained using the data corresponding to all participants except one and then tested over the left-out participant. Such a process has been reiterated as many times as there are participants and, at each repetition, a different participant has been left out. The advantage of such a design is that it is possible to test the approach over the whole corpus at disposition while still keeping separated training and test set. The SVM has been implemented using the scikit-learn (http://scikit-learn.org/). Table 2 shows the recognition results in terms of Accuracy, Precision and Recall. According to a two-tailed Binomial Test, the SVM performs better, to a statistically significant extent, than a random classifier ($p < 0.0001$).

4.1. Reading Speed

Neuroscience suggests that some brain processes revolving around language tend to take more time in people affected by depression. In particular, it has been shown that there is an association between depression and disfunctions in several areas involved in semantic language processing, including frontal gyrus and Pre-Frontal Cortex (PFC) [31]. Furthermore, while processing the meaning of words, depression patients display slower activation of the left temporoparietal (Wernicke) area and involvement of brain regions (including right lateral PFC) not activated in the case of non-depressed people [32]. Since they need more time to assign meaning to words, it is reasonable to expect that depressed participants take more time to read the text at the core of the experiments. In other words, it is reasonable to expect that the amount of time needed to read the text can act as depression marker.

In the data used for the experiments, the average time required to read the text and its standard deviation is 54.92 ± 2.66 s and 47.38 ± 1.20 s for for depressed and control participants, respectively. Given that all participants read 185 words, this corresponds to average reading speeds of 202.1 and 234.3 words per minute, respectively, for depressed and control participants. In line with the neuroscience indications above, such differences are statistically significant ($p = 0.012$ according to a two-tailed t-test). This suggests that the use of the reading time as a feature, in addition to the 32 features of the standard set, should lead to an improvement of the performance. The results of such an intervention are summarized in Table 3. Compared to the baseline classifier, the accuracy increases by 9.1 points (corresponding to a reduction of the error rate by 28.6%). According to a two-tailed binomial test, such a difference is statistically significant ($p < 0.05$). In other words, the time someone needs to read a certain text appears to act effectively as a depression marker.

4.2. Effect of Silences

The previous section shows that brain-level differences between depressed and non-depressed individuals lead to observable differences in the amount of time needed to read a text. Furthermore, the differences are consistent enough to signifcantly improve the accuracy of a baseline classifier through the addition of just one feature to the original set of 32. This section shows that it is possible to further improve the performance of the baseline classifier by taking into account how depression changes the neural processes responsible for verbal fluency. In fact, the literature shows that, in the brain of depressed people, median prefrontal cortex and angular gyrus generate interferences that result into speech disfluency, typically through the recruitment of a larger number of brain areas involved in speech initiation and higher-order language processes [33]. Not surprisingly, many studies found that depressed individuals have deficits in phonemic and verbal fluency [34]. Furthermore, improved fluency is typically used as a signal of amelioration during depression treatment [34].

One possible effect of the differences above is that depressed people tend to display more frequently intervals of time during which there is no emission of voice. Furthermore, for depressed people, these intervals of time might tend to be longer. For this reason, the voicing probability in the baseline feature set has been used to identify sequences of consecutive frames in which voice is unlikely to be emitted. This has led to the estimate of the probability $p(r)$ of r consecutive frames to show a null voicing probability. Correspondingly, it led to the identification of a minimum threshold value r_0 such that the following holds:

$$p(r \geq r_0) = \sum_{r=r_0}^{r_{max}} p(r) \leq 0.05, \quad (1)$$

where r_{max} is the maximum value of r observed in the data.

In the experiments of this work, $r_0 = 56$, corresponding to a length of 0.575 s, not far from the conventional threshold of 0.5 s used to identify pauses in linguistics. Hereafter, sequences of null voicing probability at least r_0 frames long are referred to as *silences* (Figure 1 shows the distribution of number and average length across participants). The average number of silences for depressed and control participants is 9.7 and 4.2, respectively ($p < 0.01$ according to a two-tailed t-test). When it comes to the total length, it is 11.2 s and 3.8 s for the two groups ($p < 0.01$ according to a two-tailed t-test). This sug-

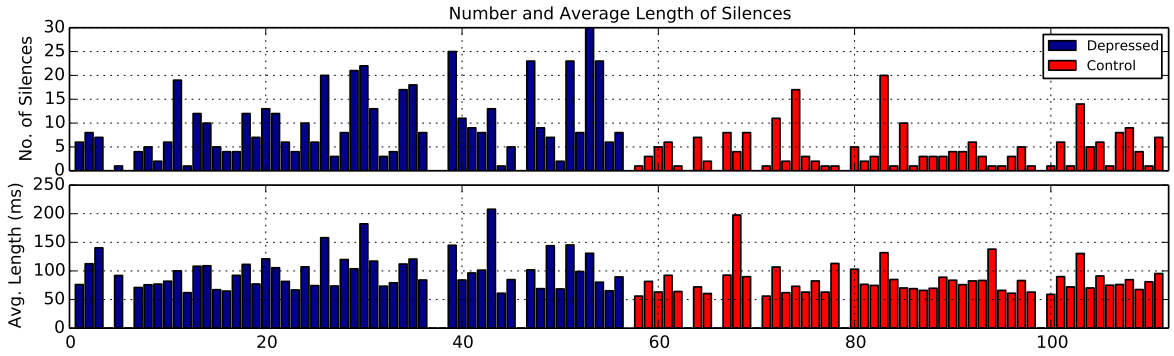

Figure 1: *The upper chart shows the number of silences per participant, the lower one shows the average length per participant.*

Table 4: *Performance after taking silences into account.*

Classifier	Features	Accuracy	Precision	Recall
Silences	35	84.5%	84.5%	84.6%

gests that the feature set can be expanded with number and total length of silences. In other words, these two features might act as depressiona markers.

The recognition results are reported in Table 4. Compared to the results obtained after adding the length of the recordings to the original set of 32 features, there is a further increase by 7.2 points of the accuracy, corresponding to a further reduction by 31.7% of the error rate (statistically significant with $p < 0.05$ according to a binomial test). Besides increasing the performance of the approach, the two features described in this section provide a possible explanation of why depressed participants tend to take more time to read the text at the core of the experiments. In particular, the effectiveness of the two features suggests that depressed individuals do not just read slower, they tend to spend more time without uttering the words they read, possibly because they need more time to process the linguistic information involved.

5. Conclusions

This article has presented experiments aimed at the identification of depression markers in speech, i.e., of measurable speech characteristics that can help to distinguish between depressed and non-depressed individuals. Compared to most previous works in the literature, the methodology used to identify the markers is not based on statistical testing or correlational analysis, but on the performance improvement observed when using the markers as features in a classifier. In particular, the experiments show that three features accounting for three different markers (reading speed, number of silences and total length of silences) increase the accuracy from 68.2% to 84.5% when added to an initial set of 32 features (statistically significant with $p < 0.0001$ according to a two-tailed binomial test). These latter were selected as a baseline because, while originally designed to capture emotion [15], are known to effectively account for a much wider spectrum of psychological phenomena, including depression (see Section 4).

In addition, compared to most previous work in the literature, this article has tried to combine computational paralinguistics [7], based on low-level speech features extracted from 25 *ms* long windows, and social signal processing [8], based on the detection of nonverbal behavioural cues associated to a phenomenon of interest. This latter aspect is important because markers should correspond to observable aspects of behaviour, given that they must be of help for the diagnosis of depression. In other words, compared to measurements like fundamental frequency or MFCCs, the advantage of observable behaviours like reading speed or use of silences is that they can be possibly observed without the need of automatic analysis.

The experiments have been performed over read speech, i.e., over recordings of people asked to read the same text. The reason behind the choice is that, in the case of spontaneous speech, markers like speed and silences can be influenced by phenomena like, e.g., the cognitive effort in planning what to say next. In other words, the use of read speech limits the effect of variability sources not necessarily related to depression. This is one of the reasons why the markers appear to be in line with the indications of neuroscience showing the depressed people tend to take more time to process linguistic information and to be more disfluent.

One interesting aspect of the markers considered in the work is that they are likely to be *honest* [35], i.e., sufficiently difficult to control consciously to allow one to fake them. For example, the silence length differences between depressed and control participants correspond to an average silence length of 1.15 *s* an 0.915 *s*, respectively. Similarly, the speed differences correspond to an average time per read word of 296 *ms* and 256 *ms* for depressed and control participants, respectively. Both differences are too subtle to be consciously controlled. Therefore, it is unlikely that a depression patient can try to look like a non-depressed individual. This is an important advantage because people affected by mental health issues can try to hide their condition in order to escape treatment, mainly to avoid the stigma associated to psychiatric problems. In this respect, the approach proposed in this work promises to be of help for clinicians dealing with potential depression patients.

6. Acknowledgements

Vinciarelli is supported by United Kingdom Research and Innovation through the *UKRI Centre for Doctoral Training in Socially Intelligent Artificial Agents* (EP/S02266X/1) and by the Engineering and Physical Sciences Research Council (EPSRC) through the grant *Socially Competent Robots* (EP/N035305/1).

7. References

[1] R. Kessler, P. Berglund, O. Demler, R. Jin, D. Koretz, K. Merikangas, A. Rush, E. Walters, and P. Wang, "The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R)," *Journal of the American Medical Association*, vol. 289, no. 23, pp. 3095–3105, 2003.

[2] R. Kessler, K. McGonagle, S. Zhao, C. Nelson, M. Hughes, S. Eshleman, H.-U. Wittchen, and K. Kendler, "Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey," *Archives of General Psychiatry*, vol. 51, no. 1, pp. 8–19, 1994.

[3] H. Gotlib and C. Hammen, *Handbook of Depression*. Guilford Press, 2008.

[4] J.-A. Tanner, J. Hensel, P. Davies, L. Brown, B. Dechairo, and B. Mulsant, "Economic burden of depression and associated resource use in Manitoba, canada," *The Canadian Journal of Psychiatry*, vol. 65, no. 5, pp. 338–346, 2019.

[5] P. Greenberg, L. Stiglin, S. Finkelstein, and E. Berndt, "The economic burden of depression in 1990." *The Journal of Clinical Psychiatry*, vol. 54, no. 11, pp. 405–418, 1993.

[6] A. Mitchell, A. Vaze, and S. Rao, "Clinical diagnosis of depression in primary care: a meta-analysis," *The Lancet*, vol. 374, no. 9690, pp. 609 – 619, 2009.

[7] B. Schuller and A. Batliner, *Computational paralinguistics: emotion, affect and personality in speech and language processing*. John Wiley & Sons, 2013.

[8] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signal Processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[9] B.-H. Lee, H. Kim, S.-H. Park, and Y.-K. Kim, "Decreased plasma BDNF level in depressive patients," *Journal of affective disorders*, vol. 101, no. 1-3, pp. 239–244, 2007.

[10] G. Placidi, M. Oquendo, K. Malone, Y.-Y. Huang, S. Ellis, and J. Mann, "Aggressivity, suicide attempts, and depression: relationship to cerebrospinal fluid monoamine metabolite levels," *Biological Psychiatry*, vol. 50, no. 10, pp. 783–791, 2001.

[11] H. Yang, C. Cusin, and M. Fava, "Is there a placebo problem in antidepressant trials?" *Current Topics in Medicinal Chemistry*, vol. 5, no. 11, pp. 1077–1086, 2005.

[12] M. Fava, A. Evins, D. Dorer, and D. Schoenfeld, "The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach," *Psychotherapy and Psychosomatics*, vol. 72, no. 3, pp. 115–127, 2003.

[13] M. Demitrack, D. Faries, D. De Brota, and W. Potter, "The problem of measurement error in multisite clinical trials," *Psychopharmacology Bulletin*, vol. 33, no. 3, pp. 513–513, 1997.

[14] R. Greenberg, R. Bornstein, M. Greenberg, and S. Fisher, "A meta-analysis of antidepressant outcome under "blinder" conditions," *Journal of Consulting and Clinical Psychology*, vol. 60, no. 5, p. 664, 1992.

[15] B. Schuller, S. Steidl, and A. Batliner, "The Interspeech 2009 emotion challenge," in *Proceedings of Interspeech*, 2009.

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

[17] M. Morales and R. Levitan, "Speech vs. text: A comparative analysis of features for depression detection systems," in *proceedings of the IEEE Spoken Language Technology Workshop*, 2016, pp. 136–143.

[18] J. Williamson, E. Godoy, M. Cha, A. Schwarzentruber, P. Khorrami, Y. Gwon, H.-T. Kung, C. Dagli, and T. Quatieri, "Detecting depression using vocal, facial and semantic communication cues," in *Proceedings of the International Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 11–18. [Online]. Available: https://doi.org/10.1145/2988257.2988263

[19] M. Rohanian, J. Hough, and M. Purver, "Detecting depression with word-level multimodal fusion," in *Proceedings of Interspeech*, 2019, pp. 1443–1447.

[20] T. Alhanai, M. Ghassemi, and J. Glass, "Detecting depression with audio/text sequence modeling of interviews," in *Proceedings of Interspeech*, 2018.

[21] Z. Huang, J. Epps, D. Joachim, and M. Chen, "Depression detection from short utterances via diverse smartphones in natural environmental conditions." in *Proceedings of Interspeech*, 2018, pp. 3393–3397.

[22] L. A. Low, N. C. Maddage, M. Lech, L. Sheeber, and N. Allen, "Detection of clinical depression in adolescents' speech during family interactions," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 574–586, 2011.

[23] N. Cummins, V. Sethu, J. Epps, S. Schnieder, and J. Krajewski, "Analysis of acoustic space variability in speech affected by depression," *Speech Communication*, vol. 75, pp. 27–49, 2015.

[24] L. Andrade, J. Caraveo-Anduaga, P. Berglund, R. Bijl, R. De Graaf, W. Vollebergh, E. Dragomirecka, R. Kohn, M. Keller, R. Kessler, N. Kawakami, C. Kiliç, D. Offord, T. Bedirhan Ustun, and H.-U. Wittchen, "The epidemiology of major depressive episodes: Results from the international consortium of psychiatric epidemiology (ICPE) surveys," *International Journal of Methods in Psychiatric Research*, vol. 12, no. 1, pp. 3–21, 2003.

[25] D. France, R. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 829–837, 2000.

[26] A. Ozdas, R. Shiavi, S. Silverman, M. Silverman, and D. Wilkes, "Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, pp. 1530–1540, 2004.

[27] T. Quatieri and N. Malyska, "Vocal-source biomarkers for depression: A link to psychomotor activity," in *Proceedings of Interspeech*, 2012.

[28] A. Nilsonne and J. Sundberg, "Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples," *Music Perception*, pp. 507–516, 1985.

[29] C. Gobl and A. Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, no. 1-2, pp. 189–212, 2003.

[30] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in OpenSMILE, the Munich open-source multimedia feature extractor," in *Proceedings of the ACM International Conference on Multimedia*, 2013, pp. 835–838.

[31] M. Seghier, F. Lazeyras, A. Pegna, J.-M. Annoni, I. Zimine, E. Mayer, C. Michel, and A. Khateb, "Variability of fMRI activation during a phonological and semantic language task in healthy subjects," *Human Brain Mapping*, vol. 23, no. 3, pp. 140–155, 2004.

[32] Y. Abdullaev, B. Kennedy, and A. Tasman, "Changes in neural circuitry of language before and after treatment of major depression," *Human Brain Mapping*, vol. 17, no. 3, pp. 156–167, 2002.

[33] H. Backes, B. Dietsche, A. Nagels, M. Stratmann, C. Konrad, T. Kircher, and A. Krug, "Increased neural activity during overt and continuous semantic verbal fluency in major depression: mainly a failure to deactivate," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 264, no. 7, pp. 631–645, 2014.

[34] S. Wagner, B. Doering, I. Helmreich, K. Lieb, and A. Tadić, "A meta-analysis of executive dysfunctions in unipolar major depressive disorder without psychotic symptoms and their changes during antidepressant treatment," *Acta Psychiatrica Scandinavica*, vol. 125, no. 4, pp. 281–292, 2012.

[35] A. Pentland, *Honest Signals*. MIT Press, 2007.