Interspeech 2024
1-5 September 2024, Kos, Greece

Exploring Gender-Specific Speech Patterns in Automatic Suicide Risk Assessment

Maurice Gerczuk[1], *Shahin Amiriparian*[2], *Justina Lutz*[3], *Wolfgang Strube*[3], *Irina Papazova*[3], *Alkomiet Hasan*[3,4], *Björn W. Schuller*[1,2,5]

[1]Chair of Embedded Intelligence for Health Care & Wellbeing, University of Augsburg, Germany
[2]CHI – Chair of Health Informatics, MRI, TU Munich, Germany
[3]District Hospital Augsburg, Germany
[4]German Center for Mental Health, Munich, Germany
[5]GLAM – Group on Language, Audio, & Music, Imperial College, UK

maurice.gerczuk@uni-a.de

Abstract

In emergency medicine, timely intervention for patients at risk of suicide is often hindered by delayed access to specialised psychiatric care. To bridge this gap, we introduce a speech-based approach for automatic suicide risk assessment. Our study involves a novel dataset comprising speech recordings of 20 patients who read neutral texts. We extract four speech representations encompassing interpretable and deep features. Further, we explore the impact of gender-based modelling and phrase-level normalisation. By applying gender-exclusive modelling, features extracted from an emotion fine-tuned wav2vec2.0 model can be utilised to discriminate high- from low suicide risk with a balanced accuracy of 81 %. Finally, our analysis reveals a discrepancy in the relationship of speech characteristics and suicide risk between female and male subjects. For men in our dataset, suicide risk increases together with agitation while voice characteristics of female subjects point the other way.

Index Terms: suicidality, computational paralinguistics, digital health

1. Introduction

Every year, more than 700 000 people die by suicide, accounting for 1 in every 100 deaths. For young people, suicide is the fourth leading cause of death following road injury, tuberculosis, and interpersonal violence [1]. Main factors contributing to suicidality can be economic pressure, past mental health issues, and personal or external crises, such as the COVID-19 pandemic which lead to an increase in suicidal ideation [2].

There are gender-specific differences in the frequency of suicidal behaviour and suicides. Women are more likely to exhibit suicidal behaviour, while suicide is more common in men. This phenomenon, known as the gender paradox [3], is in part explained by the different preferences of men and women regarding the methods of suicide they choose. Overall, men choose more lethal suicide methods like firearms or hanging [4], also referred to as "violent" suicide methods [5]. However, it also appears that the rate of completion among men is higher even when using the same suicide methods as women [6]. Freeman *et al.* [7] found that the intent to die when committing a suicide attempt was greater among men than women resulting in more serious suicide attempts vs suicide attempts that are being made as a cry for help or a means of self-harm. Evaluating data from 78 countries over 11 years, Milner *et al.* [8] showed that whilst rising gender equality led to reduced suicide rates in women, there was no significant reduction in suicide rates amongst males. The assumption of several roles in an equal social system was seen as having positive effects on female mental health whilst masculine norms linked with increased vulnerability for suicidal ideation are not being overcome at the same pace.

As professional psychiatric assessment of suicide risk is costly and often not available in emergency medicine, predictors which are easy and quick to collect are highly valuable [9]. However, a meta-study showed that traditional analysis of individual risk factors through questionnaires is of limited accuracy when predicting elevated suicide-risk [10]. In this context, biomarkers can provide a promising source of information, complementing the standard practice of screening for suicide attempt history [11]. As a biosignal that is both easy to collect as well as connected to both the physiological and cognitive systems, speech can be harnessed to analyse a wide range of health conditions [12]. For depression and suicidality, both traditional statistical analysis and machine learning strategies can be utilised to automatically infer aspects of a person's current mental health from various linguistic and paralinguistic characteristics [13].

Min *et al.* [14] fused voice characteristics with demographic information and the number of suicide attempts to classify high- against low-suicide risk and worsening of suicidality with accuracies of 69 % and 79 %, respectively. Belouali *et al.* [15] utilised acoustic and linguistic features of speech recorded from army veterans in order to automatically assess suicidal ideation using Machine Learning (ML). Based on an acoustic feature analysis, they found the voices of suicidal subjects in their exclusively male cohort to be more monotonous, dull and breathy, a finding in line with voice characteristics of suicidal adolescents [16]. Further, a person's speech fluency has been found to be a discriminative feature, as it is negatively impacted by suicidal ideation and suicide attempts [17].

Despite the differences in suicidal behaviour between women and men, few research directly targets an investigation into gender-specific correlates of speech characteristics and suicide risk. However, for the related subject of depression recognition from speech, Vlasenko *et al.* [18] found there to be a significant difference in the manifestation of depression in the formants of vowels between genders, leading to improved classification performance when utilising gender-specific features. Furthermore, Oureshi *et al.* [19] showed that adversarially learning to predict depression scores separated by gender leads to a more accurate estimation of depression severity on

10.21437/Interspeech.2024-1097

the DAIC-WOZ corpus.

In the presented work, we want to extend previous research on automatic suicide risk assessment from speech by (1) applying state-of-the-art transformer-based speech representations and comparing their efficacy against traditional audio functionals; and (2) exploring differences in speech patterns between female and male subjects via gender-based modelling and manual acoustic analysis.

2. Dataset

Our database includes speech recorded from 10 women (age $18-61$ years, $\mu = 41.7 \pm 15.5$) and 10 men (age $18-64$ years, $\mu = 37.7 \pm 16.3$) undergoing emergency admission to the psychiatric department of the District Hospital Augsburg, Germany[1]. Due to the setting and targeted application, the dataset does not include samples recorded from healthy controls, rather, every subject is diagnosed with at least one behavioural or mental disorder according to ICD-10. A majority ($N = 15$) were experiencing a severe depressive episode without psychotic symptoms which mostly occurred as part of recurrent depressive disorder (F33.2, $N = 13$). Other main diagnoses were paranoid schizophrenia (F20.0, $N = 1$), acute polymorphic psychotic disorder with ($F23.1, N = 1$) and without symptoms of schizophrenia ($F23.0, N = 1$), post-traumatic stress disorder (F43.1, $N = 1$), and adjustment disorder (F43.2, $N = 1$). The study doctor assessed the near-term suicide risk of each participant on a Likert scale of [1 – 6] (1–2: no risk, 3–4: suicidal without intention to act, 5–6: high suicide risk with intention to act). In this paper, we target a binary classification of high, near-term suicide risk (5–6) against lower suicidality (1–4). Applying this cutoff, 7 subjects – 4 male and 3 female – were at high risk of suicide.Three types of speech were obtained from every subject using a Zoom Q8 recorder: (1) two repetitions of readings of neutral texts[2]; (2) a spontaneous description of a comic; and (3) isolated vowel production. In this study, we focus our analysis on the neutral texts.

3. Methodology

Our approach begins with a preprocessing step where we normalise the volume of the audio samples and segment them into phrases via forced alignment. Afterwards, we extract both interpretable audio functionals and deep features from the segments and apply either global or phrase-level normalisation. We train and evaluate classifiers in a Leave-One-Speaker-Out (LOSO) cross-validation (CV) and investigate the effects of gender-based modelling.

3.1. Segmentation

After applying loudness normalisation, we break down the recordings of neutral texts into individual sentences. We apply forced alignment of the texts to the audio signals via whisperx [20] and split the recordings at the detected sentence boundaries. After segmentation, we receive 6, 7, and 16 segments for each recording of the three stories respectively, resulting in 1 160 speech samples across two readings of all three stories by every participant with a mean duration of 5.06 s (± 3.29 s).

[1]Ethical approval for the study was obtained from the ethics commission of the University of Augsburg.

[2]Three German short stories "Der Nordwind und die Sonne", "Gleich am Walde", and "Der Hund und das Stück Fleisch".

3.2. Feature extraction

After segmentation, we extract audio representations from each phrase separately. We follow two general paradigms of paralinguistic speech analysis by computing both interpretable sets of audio functionals as well as generating deep feature embeddings from pre-trained large-scale speech recognition models based on wav2vec2.0 (w2v2).

Audio Functionals: We extract two sets of interpretable audio features in the form of openSMILE [21] functionals – the handcrafted extended Geneva minimalistc acoustic parameter set (eGeMAPS) [22] set ($N = 88$) and the larger ComParE2016 ($N = 6,373$) feature set. Both include a range of statistical functions (e. g., mean, minimum, maximum, standard deviation, etc.) applied to low-level descriptors (LLDs) computed over consecutive short frames of the audio signal. The LLDs include parameters related to frequency (e. g., F0, jitter, centres and bandwidths of formats 1-3), energy (e. g., loudness, shimmer and harmonic to noise ratio), spectral balance and shape (e. g., Alpha Ratio, Hammarberg Index, MFCCs and spectral slopes), and temporal features (e. g., durations and frequencies of voiced segments or loudness peaks).

w2v2 embeddings: We further utilise the mean of the hidden states of the last encoder layer in pre-trained w2v2 models. We specifically select two pre-trained transformer models: wav2vec-large [23] and a 12-layer w2v2 model fine-tuned for dimensional speech emotion recognition (arousal, dominance, and valence) on MSP-Podcast [24, 25]. The former allows us to analyse whether features learnt in a self-supervised manner large-scale multi-lingual speech data capture paralinguistic markers of suicidality while the latter might be able to exploit correlations between a subject's emotional state and suicide risk.

3.3. Feature Normalisation

We apply conventional *global* feature normalisation by scaling every feature to zero mean and unit variance across the respective training datasets. Furthermore, we exploit the fixed-content nature of the collected speech samples by scaling the features on the *phrase* level. Specifically, we define the phrases as the individual sentences in the three neutral texts. We scale each sample according to statistics computed across all samples of the same sentence.

3.4. Suicide Risk Classification

All experiments utilise Support Vector Machines (SVMs) with linear kernel to classify individual phrases (either sentences or vowels). We tune the SVM's cost parameter on a logarithmic scale between 10^{-2} and 10^{-7} and balance the weights of high and low suicide risk samples during training based on their frequency.

Gender-Based Modelling: We investigate how gender-related differences in voice characteristics affect suicidality recognition performance. For this purpose, we evaluate two strategies for building gender-specific classification systems. In the first case, each model's parameters and normalisation statistics are learnt exclusively on samples of the respective gender, effectively partitioning the data before applying the pipeline. Secondly, we experiment with "soft" gender-based modelling by using weighted instance learning. Here, instead of excluding all samples of the other gender(s) from each model's training data, we apply a down-weighting factor $\lambda = 0.1$ to every out-of-group instance.

Evaluation Strategy: Due to the small size of our database,

we utilise LOSO CV. Normalisation statistics are computed on each fold's training data and hyperparameters are optimised via nested CV (5 inner folds). For model evaluation and hyperparameter tuning, we choose balanced accuracy – the mean of per class recalls – as a metric accounting for class imbalances. As the models are trained to classify suicidality per phrase but ground truth labels were assigned per subject, we present both segment-level and majority-voted subject metrics. Additionally, we report $95\,\%$ confidence intervals (CIs) computed over $1\,000$ bootstrapped samples of the model predictions. Note that bootstrapping is applied before majority voting, sometimes resulting in the actual subject level metric lying at the edge of the CI.

4. Experiments and Results

In the following, we first discuss the results of our machine learning experiments for different sets of audio features, normalisation strategies and gender-based modelling. Afterwards, we perform a feature analysis to identify the most important acoustic markers for high suicide risk.

4.1. Classification Results

Table 1 shows the results achieved for binary suicide risk classification utilising the four evaluated audio features and SVMs measured in balanced accuracy. Furthermore, global modelling is compared against two variations of gender-based modelling and features are normalised either globally or per phrase. Generally, we observe a positive impact of gender-based modelling on classification accuracy for every set of audio features, except eGeMAPS which further performs the weakest overall, only reaching a maximum $60\,\%$ segment-level and $63\,\%$ speaker-level balanced accuracy. The benefits of training gender-specific models become apparent in the larger, brute-force ComParE2016 feature set. While by itself, the models trained on these features rank below eGeMAPS, possibly due to overfitting, separating by gender helps reduce the feature space size, minimising the impact of acoustic information unrelated to suicide risk. Moreover, down-weighting ($\lambda = 0.1$) the opposite gender samples instead of completely removing them from the training data ($\lambda = 0$) increases per speaker accuracies up to $70\,\%$ while sample-level performance stays the same, indicating robustness towards inter-subject variations and increased generalisation capabilities. Finally, neither eGeMAPS nor ComParE2016 benefit from applying input normalisation per phrase instead of globally, indicating that models can easily abstract from the influence of linguistic content on these feature sets. This is further in line with findings for speech emotion recognition, where phonetic variations only have a small impact on model performance when utilising spectral features [26].

However, the results achieved with models trained on w2v2 embeddings paint a slightly different picture. Here, likely due to overfitting, baseline performance without either phrase normalisation or gender-based modelling is quite poor for features extracted from both the large cross-lingual and the emotion fine-tuned w2v2. In the case of the large XLSR model, applying phrase normalisation consistently improves per-speaker accuracy. As pointed out by Wagner *et al.* [25], w2v2 models implicitly capture linguistic information, which, however, is irrelevant to the task at hand, considering the fixed content of the neutral texts we base our analysis on. Phrase-level normalisation might help remove this confounding factor at the beginning of the machine learning pipeline. Gender-based modelling further helps performance, with the soft, down-weighting variant

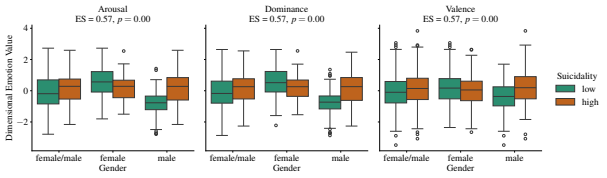

Figure 1: *Distribution of normalised arousal, dominance, and valence predictions generated by the pre-trained w2v2 speech emotion recognition model [25] in subjects with high and low suicidal risk additionally separated by gender.*

($\lambda = 0.1$) slightly edging out gender-exclusive models with a best balanced accuracy of $74\,\%$.

Interestingly, the best overall classification accuracy is achieved with the emotion-finetuned w2v2 features, but only under gender-exclusive modelling. The fact that neither phrase normalisation nor soft gender-based modelling can help models exceed near chance-level performance, could hint towards a relationship between affective states and suicidality that is specific to gender. In order to investigate this hypothesis, we extract dimensional arousal, dominance, and valence scores for every audio sample via the same fine-tuned w2v2 model used to generate the feature embeddings. We visualise the distributions of the normalised values in Figure 1. For high-risk male subjects, all three dimensions, but especially arousal, are distinctly higher than for low-risk men. Contrarily, this trend seems reversed when we look at the female subjects where higher values in the dimensions are associated with lower suicide risk. In line with these observations, Bryan *et al.* [27] found that increased agitation (a high arousal affective state) was significantly associated with the history of attempted suicide in men but did not impact suicidality for women in the same study.

4.2. Acoustic Feature Analysis

To build on these observations, we now investigate the relationship between paralinguistic speech characteristics and suicide risk – and how this relation differs across the two genders present in our dataset – by analysing a selection of acoustic features. In order to determine the most relevant features, we rank the functionals contained in eGeMAPS according to the mean absolute coefficients (feature weights) assigned to them during the training of the linear SVM models (configuration in the first column of Table 1). After ranking, we choose the top 5 functionals, sorting out redundant features, e. g., we remove the arithmetic mean of F0 as it is highly related to the 80th percentile of F0 which has a higher ranking. For each feature, we compute the non-parametric two-sided Mann-Whitney U test [28] comparing the globally normalised feature distributions between low- and high-risk subjects. These distributions, additionally disaggregated by gender, and the corresponding common language effect sizes – the proportion of low and high-risk feature pairs where the value of the high-risk sample is greater than that of the low-risk sample – and p-values are visualised in Figure 2.

$\mathbf{SlopeV0-500\,Hz_{mean}}$: We start with the linear regression slope fitted to the logarithmic power spectrum of voiced frames in the 0-500 Hz band which shows a negative correlation with suicide risk for women but only a negligible connection to suicidality for male subjects. The spectral slope has previously been shown to decrease with negative, low arousal emotions [29] and is further negatively affected by depression [30, 31]. Moreover,

Balanced Accuracy [%]	Global Modelling		Gender-based Modelling			
			$\lambda = 0$		$\lambda = 0.1$	
Features	global normalisation	phrase normalisation	global normalisation	phrase normalisation	global normalisation	phrase normalisation
ComParE2016	52 (49-55) / 59 (45-63)	55 (52-57) / 59 (48-66)	62 (59-65) / 63 (55-70)	61 (59-64) / 59 (59-70)	61 (57-63) / 70 (63-78)	64 (61-67) / 70 (66-78)
eGeMAPSv02	58 (55-61) / 63 (49-70)	54 (51-57) / 46 (38-53)	56 (53-59) / 60 (60-64)	52 (49-55) / 60 (46-67)	60 (57-63) / 60 (60-64)	54 (51-57) / 49 (46-57)
wav2vec2-audeering-emo-dim	57 (54-60) / 49 (45-56)	54 (51-56) / 45 (41-52)	66 (63-69) / 74 (67-82)	**70 (68-73) / 81 (70-85)**	54 (52-57) / 49 (38-56)	58 (55-61) / 52 (48-60)
wav2vec2-large-xlsr-53	52 (49-55) / 53 (49-60)	55 (52-58) / 65 (52-69)	60 (57-63) / 56 (49-67)	59 (56-63) / 67 (56-74)	58 (55-61) / 67 (56-71)	61 (58-64) / 74 (60-74)

Table 1: *Classification results for audio-based suicide risk assessment measured in balanced accuracy computed sample-wise and additionally aggregated per speaker by majority vote (after forward slash). 95 % confidence intervals computed from bootstrapped model predictions are given in parentheses. Both global – one classifier is trained to generate predictions for all samples – and gender-based modelling is evaluated. For gender-based modelling, each gender's classifier is trained via weighted instance learning, applying a down-weighting factor λ to all samples of the other gender(s). In our dataset, only women and men are represented.*

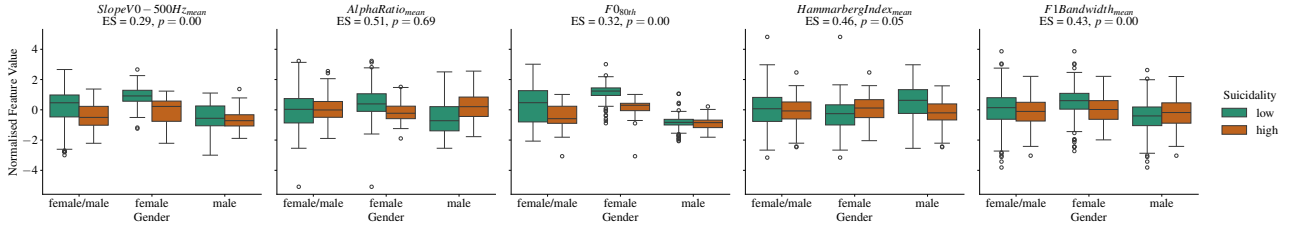

Figure 2: *Distribution of most important features determined by effect size of Mann-Whitney U test for low and high suicidality. Additionally split by gender of speaker.*

Ozdas *et al.* [32] found that the spectral slope decreases between healthy controls and both depressed and near-term suicidal individuals, but patients with higher suicide risk exhibit increased glottal spectral slopes compared to depressed, non-suicidal subjects. However, their analysis was only based on male subjects.

$\mathbf{AlphaRatio_{mean}}$: Analysed across the whole dataset, there is no relationship between the ratio of spectral energy above and below 1 000 Hz and suicidality. However, when looking at the genders separately, a shift of spectral energy towards 1 kHz – 5 kHz (higher alpha ratio) is associated with an increased suicide risk for male subjects while the opposite relationship can be observed for women. Higher alpha ratio can be a product of increased emotional arousal or activity [33], aligning with the behaviour of the emotion fine-tuned w2v2 outlined in Section 4.1

$\mathbf{HammarbergIndex_{mean}}$: Defined as the difference between the maximum energy peak in the 0-2 kHz and the 2-5 kHz bands, the Hammarberg Index is considered an indicator of vocal effort [34]. Like the alpha ratio, an energy increase in the upper-frequency bands (lower Hammarberg Index) is often associated with higher arousal and basic emotions such as hot anger [35]. Accordingly, we observe a decrease in the Index with higher suicide risk in male subjects, while a less pronounced trend in the opposite direction can be found for women.

$\mathbf{F0_{80th}}$: While no difference in mean F0 between low- and high-risk men exists in our dataset, women in the high-risk group display significantly reduced pitch, a feature that is often correlated with depression severity [13].

$\mathbf{F1Bandwidth_{mean}}$: Finally, the bandwidth of the first formant narrows with higher suicide risk for female subjects, indicating less breathiness. Contrary to the trends observed in the other analysed features, this is somewhat atypical for speech produced under increased depressive symptom severity [36].

Overall, the acoustic feature analysis further supports our findings with the emotion fine-tuned w2v2 model. Specifically, suicidality is reflected differently in the voices of women and men in our dataset. While high-risk female subjects exhibit speech patterns that align with an increase in depressive symptomatology (e. g., decreased pitch and narrowed spread of spectral energy), higher suicide risk in men comes with paralinguistic characteristics that are associated with heightened activation and emotional arousal.

5. Conclusions and Future Work

In the present work, we utilised a novel database for the automatic assessment of suicide risk from speech in emergency medicine and investigated gender-specific correlations between suicidality and paralinguistics. Our machine learning results and a consecutive acoustic feature analysis indicate that gender-based modelling can be effective for discriminating between low- and high-risk individuals. In our data, high suicidality in men positively correlates with paralinguistic characteristics that are usually associated with a high-arousal, agitated affective state while the speech patterns of female subjects point in the opposite direction. This finding could be connected to existing literature on suicide risk [27] which shows a male-exclusive correlation of agitation and suicide attempt history. The results further indicate the efficacy of deep, transformer-based speech representations, even in a small database. A **limitation** of our study can be found with the small size of our database, containing only 20 subjects – which, however, is not unusual in the domain. **Future work** should extend the gender-based analysis and modelling to other types of voice recordings, e. g.vowel productions and unscripted, variable content speech on larger databases.

6. Acknowledgements

This work was supported by MDSI – Munich Data Science Institute as well as MCML – Munich Center of Machine Learning. B.W.S is also with the Konrad Zuse School of Excellence in Reliable AI (relAI) in Munich, Germany.

7. References

[1] Geneva: World Health Organization, "Suicide worldwide in 2019: Global health estimates," 2021.

[2] Y. Yan, J. Hou, Q. Li, and N. X. Yu, "Suicide before and during the COVID-19 pandemic: A systematic review with meta-analysis," *Int. J. Environ. Res. Public. Health*, vol. 20, no. 4, p. 3346, Feb. 2023.

[3] S. S. Canetto and I. Sakinofsky, "The gender paradox in suicide," *Suicide Life. Threat. Behav.*, vol. 28, no. 1, pp. 1–23, 1998.

[4] V. J. Callanan and M. S. Davis, "Gender differences in suicide methods," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 47, no. 6, pp. 857–869, Jun. 2012.

[5] B. Ludwig and Y. Dwivedi, "The concept of violent suicide, its underlying trait and neurobiology: A critical perspective," *Eur. Neuropsychopharmacol.*, vol. 28, no. 2, pp. 243–251, Feb. 2018.

[6] A. Cibis, R. Mergl, A. Bramesfeld, D. Althaus, G. Niklewski, A. Schmidtke, and U. Hegerl, "Preference of lethal methods is not the only cause for higher suicide rates in males," *J. Affect. Disord.*, vol. 136, no. 1-2, pp. 9–16, Jan. 2012.

[7] A. Freeman, R. Mergl, E. Kohls, A. Székely, R. Gusmao, E. Arensman, N. Koburger, U. Hegerl, and C. Rummel-Kluge, "A cross-national study on gender differences in suicide intent," *BMC Psychiatry*, vol. 17, no. 1, p. 234, Dec. 2017.

[8] A. Milner, A. J. Scovelle, B. Hewitt, H. Maheen, L. Ruppanner, and T. L. King, "Shifts in gender equality and suicide: A panel study of changes over time in 87 countries," *J. Affect. Disord.*, vol. 276, pp. 495–500, Nov. 2020.

[9] R. Iyer and D. Meyer, "Detection of suicide risk using vocal characteristics: Systematic review," *JMIR Biomed. Eng.*, vol. 7, no. 2, e42386, Dec. 2022.

[10] J. C. Franklin, J. D. Ribeiro, K. R. Fox, K. H. Bentley, E. M. Kleiman, X. Huang, K. M. Musacchio, A. C. Jaroszewski, B. P. Chang, and M. K. Nock, "Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research.," *Psychol. Bull.*, vol. 143, no. 2, pp. 187–232, 2017.

[11] K. Sudol and J. J. Mann, "Biomarkers of suicide attempt behavior: Towards a biological model of risk," *Curr. Psychiatry Rep.*, vol. 19, no. 6, p. 31, Jun. 2017.

[12] N. Cummins, A. Baird, and B. W. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods*, vol. 151, pp. 41–54, Dec. 2018.

[13] 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. Commun.*, vol. 71, pp. 10–49, Jul. 2015.

[14] S. Min, D. Shin, S. J. Rhee, C. H. K. Park, J. H. Yang, Y. Song, *et al.*, "Acoustic analysis of speech for screening for suicide risk: Machine learning classifiers for between- and within-person evaluation of suicidality," *J. Med. Internet Res.*, vol. 25, no. 1, e45456, Mar. 2023.

[15] A. Belouali, S. Gupta, V. Sourirajan, J. Yu, N. Allen, A. Alaoui, M. A. Dutton, and M. J. Reinhard, "Acoustic and language analysis of speech for suicidal ideation among US veterans," *BioData Min.*, vol. 14, no. 1, p. 11, Feb. 2021.

[16] S. Scherer, J. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *ICASSP*, Vancouver, Canada, May 2013, pp. 709–713.

[17] B. Stasak, J. Epps, H. T. Schatten, I. W. Miller, E. M. Provost, and M. F. Armey, "Read speech voice quality and disfluency in individuals with recent suicidal ideation or suicide attempt," *Speech Commun.*, vol. 132, pp. 10–20, Sep. 2021.

[18] B. Vlasenko, H. Sagha, N. Cummins, and B. Schuller, "Implementing gender-dependent vowel-level analysis for boosting speech-based depression recognition," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden: ISCA, Aug. 2017, pp. 3266–3270.

[19] S. A. Oureshi, G. Dias, S. Saha, and M. Hasanuzzaman, "Gender-aware estimation of depression severity level in a multimodal setting," in *IJCNN*, virtual, Jul. 2021, pp. 1–8.

[20] M. Bain, J. Huh, T. Han, and A. Zisserman, "WhisperX: Time-accurate speech transcription of long-form audio," in *Proc. INTERSPEECH 2023*, Dublin, Ireland, Aug. 2023, pp. 4489–4493.

[21] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *ACM Multimedia 2010*, Firenze, Italy: ACM Press, Oct. 2010, pp. 1459–1462.

[22] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. Andre, C. Busso, *et al.*, "The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *TAFFC*, vol. 7, no. 2, pp. 190–202, Apr. 2016.

[23] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proc. NeurIPS 2020*, vol. 33, virtual: Curran Associates, Inc., 2020, pp. 12 449–12 460.

[24] R. Lotfian and C. Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," *TAFFC*, vol. 10, no. 4, pp. 471–483, Oct. 2019.

[25] J. Wagner, A. Triantafyllopoulos, H. Wierstorf, M. Schmitt, F. Burkhardt, F. Eyben, and B. W. Schuller, "Dawn of the transformer era in speech emotion recognition: Closing the valence gap," *TPAMI*, vol. 45, no. 9, pp. 10 745–10 759, Sep. 2023.

[26] V. Sethu, E. Ambikairajah, and J. Epps, "Speaker dependency of spectral features and speech production cues for automatic emotion classification," in *ICASSP 2009*, Taipei, Taiwan: IEEE, Apr. 2009, pp. 4693–4696.

[27] C. J. Bryan, M. J. Hitschfeld, B. A. Palmer, K. M. Schak, E. M. Roberge, and T. W. Lineberry, "Gender differences in the association of agitation and suicide attempts among psychiatric inpatients," *Gen. Hosp. Psychiatry*, vol. 36, no. 6, pp. 726–731, Nov. 2014.

[28] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, vol. 18, no. 1, pp. 50–60, 1947. JSTOR: 2236101.

[29] M. Guzman, S. Correa, D. Muñoz, and R. Mayerhoff, "Influence on spectral energy distribution of emotional expression," *J. Voice*, vol. 27, no. 1, 129.e1–129.e10, Jan. 2013.

[30] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An investigation of depressed speech detection: Features and normalization," in *Proc. INTERSPEECH 2011*, Firenze, Italy: ISCA, Aug. 2011, pp. 2997–3000.

[31] F. Hönig, A. Batliner, E. Nöth, S. Schnieder, and J. Krajewski, "Automatic modelling of depressed speech: Relevant features and relevance of gender," in *Proc. INTERSPEECH 2014*, Singapore: ISCA, Sep. 2014, pp. 1248–1252.

[32] 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 Trans. Biomed. Eng.*, vol. 51, no. 9, pp. 1530–1540, Sep. 2004.

[33] T. Waaramaa, A.-M. Laukkanen, M. Airas, and P. Alku, "Perception of emotional valences and activity levels from vowel segments of continuous speech," *J. Voice*, vol. 24, no. 1, pp. 30–38, Jan. 2010.

[34] B. Hammarberg, B. Fritzell, J. Gaufin, J. Sundberg, and L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta Otolaryngol.*, vol. 90, no. 1-6, pp. 441–451, Jan. 1980.

[35] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression.," *J. Pers. Soc. Psychol.*, vol. 70, no. 3, pp. 614–636, 1996.

[36] N. Cummins, J. Dineley, P. Conde, F. Matcham, S. Siddi, F. Lamers, *et al.*, *Multilingual Markers of Depression in Remotely Collected Speech Samples*. Oct. 2022.