



The 1st SpeechWellness Challenge: Detecting Suicide Risk Among Adolescents

Wen Wu^{*,1,4}, Ziyun Cui^{*,2,1}, Chang Lei^{*,3}, Yinan Duan³, Diyang Qu³
Ji Wu², Bowen Zhou^{1,2}, Runsen Chen^{†,3}, Chao Zhang^{†,1,2}

¹ Shanghai Artificial Intelligence Laboratory, China

² Department of Electronic Engineering, Tsinghua University, China

³ Vanke School of Public Health, Tsinghua University, China

⁴ Department of Engineering, University of Cambridge, UK

wuwen@pjlab.org.cn, cui-zy24@mails.tsinghua.edu.cn, cz277@tsinghua.edu.cn

Abstract

The 1st SpeechWellness Challenge (SW1) aims to advance methods for detecting current suicide risk in adolescents using speech analysis techniques. Suicide among adolescents is a critical public health issue globally. Early detection of suicidal tendencies can lead to timely intervention and potentially save lives. Traditional methods of assessment often rely on self-reporting or clinical interviews, which may not always be accessible. The SW1 challenge addresses this gap by exploring speech as a non-invasive and readily available indicator of mental health. We release the SW1 dataset which contains speech recordings from 600 adolescents aged 10-18 years. By focusing on speech generated from natural tasks, the challenge seeks to uncover patterns and markers that correlate with current suicide risk.

Index Terms: speech wellness, suicide risk detection, speech analysis

1. Introduction

Suicide remains a critical global health challenge and is among the leading causes of death among adolescents worldwide [1, 2, 3]. Early detection of suicide risk is essential for effective prevention and intervention of potential suicide attempts. The diagnosis of suicide risk is challenging, as individuals exhibit different behaviours, making it difficult to obtain a unified clinical representation of a suicidal individual. It also relies heavily on the participant's ability, willingness, and honesty in communicating their symptoms, emotions, and thoughts [4]. Clinical interviews and self-report questionnaires are commonly utilised in clinical practice for diagnosis. However, clinical interviews are resource-intensive, requiring substantial human effort and the expertise of well-trained clinicians. Questionnaires, although more cost-effective as screening tools, are prone to issues such as bias, discrimination, and deliberate disguise [5]. This motivates the study of automatic suicide risk detection [6, 7, 8, 9, 10, 11, 12].

Speech has been shown as a promising biomarker for detecting mental disorders [13, 14, 15] and cognitive diseases [16, 17, 18], as well as suicide risk [4, 19, 20]. Speech provides a rich source of both semantic and non-semantic (paralinguistic) information, and can be collected in a cheap, remote, non-invasive and non-intrusive way. From the perspective of semantics, suicidal speech tends to have different top words from non-suicidal speech [21]. From the perspective of paralinguistics, difference can be observed in spectral properties, source

features (*e.g.*, jitter, shimmer), prosodic features (*e.g.*, F0), formant characteristics [4], along with increased disfluencies, such as more frequent hesitations and speech errors [22].

In line with Interspeech 2025's conference theme of *Fair and Inclusive Speech Science and Technology*, we launch the 1st SpeechWellness challenge, with the aim of bringing attention to the importance of mental health in speech technology research and bridging the gap between speech technology, psychology, and healthcare. To the best of our knowledge, this is the first such challenge focused on suicide risk among adolescents. It is hoped that the challenge could motivate novel algorithms and models that can detect subtle cues in speech related to current suicide risk as well as providing evidence that could inform mental health policies and clinical practices.

2. Challenge Description

The 1st SpeechWellness Challenge (abbreviated as SW1) focuses on advancing techniques for detecting current suicide risk among adolescents through speech analysis. Challenge teams are tasked with developing models that utilise spontaneous and reading speech as digital biomarkers for binary current suicide risk detection (has current suicide risk or not).

The dataset used for the challenge comprises speech recordings from 600 Chinese teenagers aged 10-18 years¹. All participants (speakers) have been anonymised before making this dataset available to the challenge teams, which provides a unique opportunity to apply and refine advanced speech technologies for public health while preserving the privacy of all participants. Details about the dataset and anonymisation method will be discussed in Section 3. Challenge teams are encouraged to leverage a variety of methodologies, including but not limited to signal processing, self-supervised learning foundation models, speech recognition, emotion recognition, and large language models. The goal is to push the boundaries of current technology to create effective tools for early detection and intervention of suicide risk.

3. The SW1 Challenge Dataset

3.1. Data collection

Data was collected from 47 elementary and middle schools in Yunfu city, southern China, consisting of 600 participants aged 10 to 18 years. At-risk students were identified based on professional clinical interviews. The control group includes non-risk

¹The challenge data is shared solely for scientific research purposes after carefully anonymisation and is to be used solely for challenge purposes, with strict prohibitions on redistribution or any other use.

* Equal contribution. † Corresponding author. Supported by Shanghai Artificial Intelligence Laboratory.

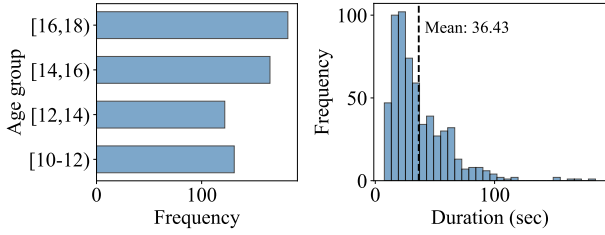


Figure 1: Distribution of age group.

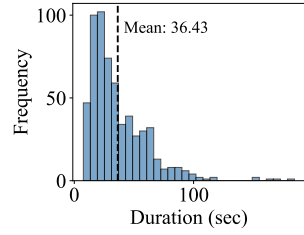


Figure 2: Histogram of recording duration of the ER task.

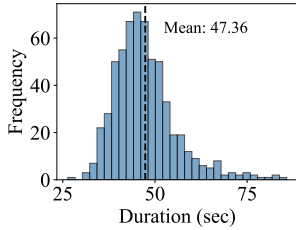


Figure 3: Histogram of recording duration of the PR task.

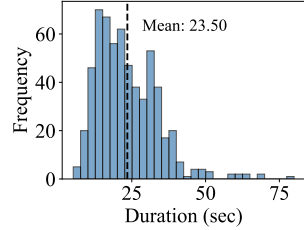


Figure 4: Histogram of recording duration of the ED task.

students. Informed consent was obtained from all participants and their guardians before data collection, with participants informed of their rights to withdraw at any time or decline uncomfortable questions. Interviews were conducted by trained interviewers specialising in related fields such as psychology and psychiatry, under the supervision of psychiatrists.

The interview contains both open-ended (unstructured tasks) and structured tasks (such as paragraph reading). Data collection was conducted in quiet, sound-proof rooms with only the interviewer and the participant present to avoid disturbances and maintain privacy. Standardised voice recorders were used by interviewers to capture speech data from participants’ speech tasks, ensuring consistency in audio format and quality. All interviews were conducted in Mandarin. The project has obtained ethical approval from the relevant committee.

The recordings used for the SW1 challenge consist of each participant’s responses to three different speech tasks designed to capture a range of linguistic, cognitive, and emotional features.

1. **Emotional Regulation (ER):** Participants are asked to answer an open-ended question, “Have you ever experienced moments of extreme emotional distress? How do you manage such feelings?”
2. **Passage Reading (PR):** Participants read the poetry *The North Wind and The Sun*, a passage from Aesop’s Fables commonly used in linguistic studies for examining structures such as morphemes [23]².
3. **Expression Description (ED):** Given an image of a facial expression [24], participants are asked to describe it³.

Diagnosis of current suicide risk was conducted by professionally trained interviewers based on a professional suicide risk assessment: the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID) suicide diagnostic [25]. It is a structured diagnostic interview designed for

²The passage can be found at: <https://github.com/speechwellness/supplementary-materials/blob/main/README.md>.

³The image can be found at: <https://github.com/speechwellness/supplementary-materials/blob/main/README.md>.

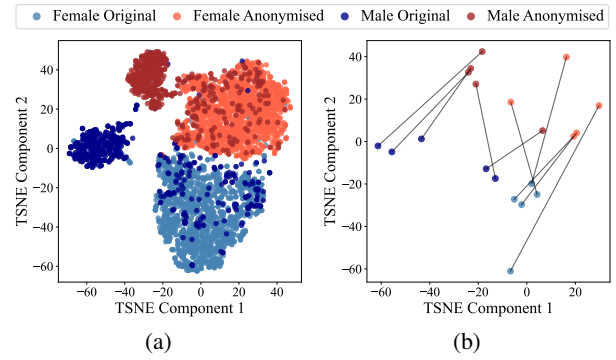


Figure 5: (a) TSNE visualisation of speaker embeddings before (blue) and after (red) anonymisation. Dark points represent male speakers, light points represent female speakers. (b) Speaker embeddings of 10 randomly selected cases before and after anonymisation, paired with lines.

adolescents, which is considered a gold standard interview for assessing adolescent psychopathology and has been validated for use by clinicians and researchers to assess current suicide risk based on DSM-IV and ICD-10 criteria [26, 27, 28].

The SW1 dataset comprises speech recordings from 600 participants (420 females and 180 males), evenly divided between 300 non-risk and 300 at-risk individuals. Statistics of age group is shown in Figure 1. The duration of speech recordings of each task is shown in Figure 2-4 with the average duration shown by dashed lines. The dataset is split into training, development (dev), and test sets in a 4:1:1 ratio. Consistency in gender distributions across the train/dev/test splits was ensured. The challenge teams are required to develop a model to predict the label (has current suicide risk or not) for a participant. They are free to choose whether to use recordings from a single task or a combination of all three tasks.

3.2. The timbre anonymisation process

Given that the participants are adolescents, a vulnerable population requiring heightened ethical considerations, we adopt additional anonymisation methods to further protect their privacy. Neural voice conversion techniques⁴ were applied to alter the timbre of the participants’ voices while preserving prosody and rhythm so that the essential features for developing the automatic suicide risk detection system remain intact.

To illustrate the impact of timbre modification, speaker embeddings were extracted from the recordings before and after voice conversion using a pretrained x-vector model⁵. These embeddings are then visualised in Figure 5(a). The blue points represent speaker embeddings before anonymisation, while the red points represent embeddings after anonymisation. The figure clearly illustrates that the clusters are distinct and separated. The smaller clusters in the upper left, shown in darker colours and set apart from the larger clusters, primarily represent male speakers, while the larger clusters predominantly correspond to female speakers. Since the participants include children in the process of voice maturation, some dark points (male) may appear within the lighter clusters (female). Figure 5(b) shows speaker embeddings of 10 randomly selected cases before and

⁴Available at: <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI>

⁵Available at: <https://huggingface.co/microsoft/wavlm-base-plus-sv>

Table 1: CER between the ground truth and the transcriptions from the original and anonymised recordings, as well as between the original and anonymised transcriptions.

Ground truth	Original	Anonymised	%CER
✓	✓		21.4
✓		✓	23.8
	✓	✓	6.0

after anonymisation, paired with lines to demonstrate the non-linear nature of the anonymisation transformation.

To quantify the impact of anonymisation on speaker verifiability, we compute the equal error rate (EER), which compares pairs of positive (same speaker) and negative (different speakers) utterances based on cosine similarity of x-vectors to verify speaker identity [29]. Post-anonymisation EER is 30.28% while the original EER is 9.33%. The increase in EER signifies enhanced anonymisation efficacy.

To further assess the impact of timbre anonymisation on speech intelligibility, the character error rate (CER) was calculated for the original and anonymised recordings⁶. We first obtain transcriptions for the original and anonymised recordings using a pretrained ASR system⁷, then compute the CER between the ground truth and the transcriptions from the original and anonymised recordings, as well as between the original and anonymised transcriptions. Results are listed in Table 1. The anonymised recordings yield a slightly higher CER than the original recordings, with a 6% CER between transcriptions before and after anonymisation, which is within an acceptable range.

3.3. Response validation and privacy check

Apart from the timbre anonymisation process, we also took action to filter out data that doesn’t meet the requirement. The validation process consists of four stages:

1. Response length: check the response contains sufficient content.
2. Task relevance: check the interviewee’s responses are relevant to the assigned task.
3. Sensitive information exposure: ensure that no sensitive privacy-related information is revealed in the response.
4. Timbre modification quality: check the intelligibility of the speech after voice conversion.

For the response length check, we filtered out recordings with transcriptions shorter than a threshold of 10 words. To assess the speech quality after timbre modification, we set a CER threshold of 40% between the transcriptions of the original and modified speech and filtered out recordings with a CER exceeding this threshold. We employed GPT-4o⁸ to check task relevance and sensitive information exposure. The instructions were presented in a few-shot manner. Along with the prompt, we provided a positive and a negative case to help the agent better complete the task⁹.

⁶Recordings of the passage reading task were used, as ground truth transcriptions were available.

⁷Available at: https://www.modelscope.cn/models/iic/speech_paraformer-large-vad-punc_asr_nat-zh-cn-16k-common-vocab8404-pytorch/summary

⁸<https://platform.openai.com/docs/models#gpt-4o>

⁹Detailed prompts can be found at <https://github.com/speechwellness/supplementary-materials/blob/>

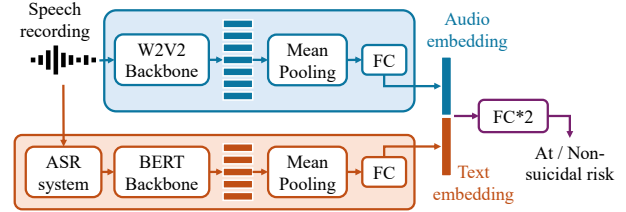


Figure 6: Illustration of the structure of the W2V2+BERT baseline. FC denotes a fully-connected layer.

4. Baseline Systems

Two baseline systems were provided for the challenge: “Baseline” system based on hand-crafted features and traditional machine learning classifiers, and “Baseline-bonus” system based on foundation models and neural network classifiers.

4.1. Baseline: eGeMAPS+SVM

The extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) is a standard set of acoustic features commonly used in speech analysis. It contains the fundamental frequency (F0), energy features (*e.g.*, loudness), spectral features (*e.g.*, formant frequencies, HNR), voice quality (*e.g.*, jitter, shimmer), as well as their most common statistical functionals, for a total of 88 features per speech segment. The eGeMAPS+SVM baseline extracts the eGeMAPS feature set from the input speech segment and feeds the features into a support vector machine (SVM) for classification. Results of the three tasks were combined by voting. The system was implemented using the scikit-learn toolkit. The ER and ED tasks used RBF kernel with $\gamma = 100$ while the PR task used linear kernel. The L2 penalty coefficient was set to 10 for all tasks. The default values were kept for all other hyperparameters.

The eGeMAPS+SVM system serves as the standard baseline, a reference point for challenge teams to compare against.

4.2. Baseline-bonus: W2V2+BERT

The overall structure of the W2V2+BERT baseline is shown in Figure 6, which contains an audio branch and a text branch. The backbone of the audio branch uses a wav2vec 2.0 (W2V2) model¹⁰ [30], which contains 24 Transformer blocks with model dimension 1024 and a total number of 317M parameters, pretrained on 56k hours of multilingual speech data. The backbone of the text branch uses a BERT model¹¹ [31], which contains 12 Transformer blocks with model dimension 768 and a total number of 110M parameters, pretrained on Chinese with a vocabulary size of 21,128.

The training consisted of two stages. In the first stage, the W2V2 and the BERT backbones were separately finetuned for suicide risk detection. In the second stage, features extracted from the two backbones were concatenated and used to train the fused suicide risk classifier. For the PR task where participants read the same paragraph, only the W2V2 branch was used. The models were trained by cross-entropy loss. The predictions from the three tasks were combined by averaging their predic-

main/README.md.

¹⁰Available at: <https://huggingface.co/facebook/wav2vec2-large-xlsr-53>

¹¹Available at: <https://huggingface.co/google-bert/bert-base-chinese>

Table 2: Training parameters for the W2V2+BERT baseline.

		Learning rate			Batch size
		ER	PR	ED	
Stage 1	Finetune W2V2	2e-5	5e-5	5e-5	8
	Finetune BERT	5e-5	N/A	5e-5	16
Stage 2	Fusion	1e-5	1e-3	5e-5	16

Table 3: Accuracy of the two baselines on the dev and test sets.

Dev / Test	eGeMAPs+SVM	W2V2+BERT
Emotional Regulation	0.52 / 0.51	0.56 / 0.54
Paragraph Reading	0.54 / 0.48	0.57 / 0.54
Expression Description	0.58 / 0.57	0.54 / 0.60
Combined	0.53 / 0.51	0.56 / 0.61

tive probabilities (softmax outputs), and the final prediction was determined as the class with higher averaged probability.

The models were implemented using PyTorch with Adam optimiser and Cosine learning rate scheduler. The learning rate and batch size used for each task and stage are listed in Table 2. The speech recordings were chunked before feeding into the W2V2 model with a chunk size of 10 s and a chunk shift of 6 s. All experiments were conducted on a single Nvidia 4090 GPU.

The W2V2+BERT system serves as the bonus baseline, an additional challenge designed for teams aiming to exceed a higher standard and earn a bonus distinction.

4.3. Results

The performance is evaluated by accuracy and the results are listed in Table 3. Overall, W2V2+BERT system outperforms eGeMAPs+SVM system. Among the three tasks, the ED task achieves slightly better results. However, the combination of the three tasks does not necessarily lead to improvement, possibly due to the limited capability of the models for each individual task. Challenge teams are encouraged to explore more advanced models and sophisticated techniques to enhance suicide risk detection performance. They may choose to use recordings from a single task or combine data from different tasks. The combination of three tasks in the W2V2+BERT system yields an accuracy of 56% on the dev set and 61% on the test set. For the original speech before timbre anonymisation, the accuracy is 58% on the dev set and 60% on the test set, which suggests that the anonymisation process preserves the information relevant to the task.

5. Summary of Participation

The SW1 challenge has received 127 registrations and 186 system submissions from teams worldwide. The score distribution is visualised in Figure 7. We are pleased to report that 68.8% of the systems surpassed the Baseline system and 14 systems beat the Baseline-bonus! The top-performing system achieved an accuracy of 0.74 on the test set, representing a 21.3% relative improvement over the Baseline-bonus. The score distribution provides insight into the relative difficulty of the task and the effectiveness of different approaches.

6. Conclusion

This paper introduces the 1st SpeechWellness Challenge (SW1), the first challenge targeting at current suicide risk de-

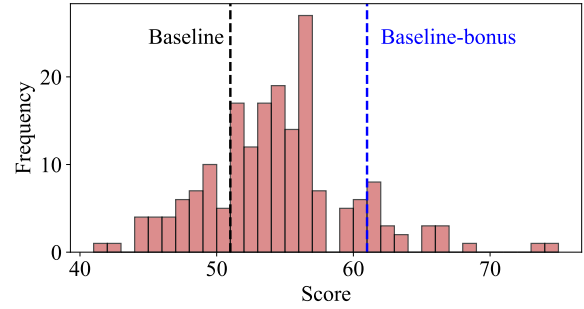


Figure 7: Distribution of scores from challenge submissions.

tection among adolescents through speech analysis. The dataset collected for the challenge comprises speech recordings from 600 adolescents, each performing three speech tasks: emotional regulation question, passage reading, and expression description. All recordings have undergone timbre modification to ensure anonymisation and protect the privacy of the participants. A further filtering process was also applied to ensure the effectiveness of the responses, and no sensitive information is disclosed. Apart from the standard Baseline system, we also introduced a Baseline-bonus system for the challenge, targeting teams aiming to exceed a higher benchmark and achieve a bonus distinction. The challenge received 127 registrations, and the best-performing system achieved an accuracy of 0.74 on the test set. The SW1 challenge seeks to highlight the importance of mental health in speech technology research, advances the development of novel algorithms to detect subtle cues in speech related to current suicide risk, and foster collaboration across the fields of speech technology, psychology, and healthcare.

7. Limitations and Ethical Consideration

This study’s findings are based on the MINI-KID scale, which assesses current suicide risk based on participants’ immediate responses. While widely recognised as a gold standard for adolescent suicide risk assessment, MINI-KID has limitations, including reliance on self-reported data, which may not fully capture the complexity of suicidal ideation. These results should not be interpreted as predictions of future suicidal behaviour and are strictly confined to this assessment’s context.

Generalisation is a common challenge in machine learning, especially in this task, where privacy concerns and the sensitive nature of the issue limit data availability and variety. The test set labels were withheld during the challenge to ensure systems were evaluated on unseen data. However, the dataset may not fully reflect the diversity of suicidal behaviours across different populations, as factors like culture and socioeconomic status could affect how generalizable the findings are.

Automatic prediction of suicide risk remains challenging, as reflected in our results, with an inherent risk of misclassification. This issue extends beyond suicide risk to medical tasks where errors have serious consequences. Developing such systems is still an ongoing process, requiring continuous refinement of models, improved data representation, and interdisciplinary collaboration – precisely why we launched this challenge. Given the complexity of this task and the need for further advancements, the SW1 challenge aims to encourage innovative solutions, foster collaboration, and benchmark approaches to advance reliable, ethical suicide risk prediction models.

8. References

- [1] World Health Organization, “Suicide worldwide in 2019: Global health estimates,” 2021.
- [2] B. Shain, P. K. Braverman, W. P. Adelman, E. M. Alderman, C. C. Breuner, D. A. Levine, A. V. Marcell, R. F. O’Brien *et al.*, “Suicide and suicide attempts in adolescents,” *Pediatrics*, vol. 138, no. 1, p. e20161420, 2016.
- [3] M. Orri, S. Scardera, L. C. Perret, D. Bolanis, C. Temcheff, J. R. Séguin, M. Boivin, G. Turecki, R. E. Tremblay, S. M. Côté *et al.*, “Mental health problems and risk of suicidal ideation and attempts in adolescents,” *Pediatrics*, vol. 146, no. 1, p. e20193823, 2020.
- [4] 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.
- [5] L. Ganzini, L. M. Denneson, N. Press, M. J. Bair, D. A. Helmer, J. Poat, and S. K. Dobscha, “Trust is the basis for effective suicide risk screening and assessment in veterans,” *Journal of general internal medicine*, vol. 28, pp. 1215–1221, 2013.
- [6] J. Oh, K. Yun, J.-H. Hwang, and J.-H. Chae, “Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales,” *Frontiers in Psychiatry*, vol. 8, p. 192, 2017.
- [7] Q. Cheng, T. M. Li, C.-L. Kwok, T. Zhu, and P. S. Yip, “Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study,” *Journal of Medical Internet Research*, vol. 19, no. 7, p. e243, 2017.
- [8] C. Su, R. Aseltine, R. Doshi, K. Chen, S. C. Rogers, and F. Wang, “Machine learning for suicide risk prediction in children and adolescents with electronic health records,” *Translational Psychiatry*, vol. 10, no. 1, p. 413, 2020.
- [9] A. Roy, K. Nikolitch, R. McGinn, S. Jinah, W. Klement, and Z. A. Kaminsky, “A machine learning approach predicts future risk to suicidal ideation from social media data,” *NPJ Digital Medicine*, vol. 3, no. 1, p. 78, 2020.
- [10] T. Zhang, A. M. Schoene, and S. Ananiadou, “Automatic identification of suicide notes with a Transformer-based deep learning model,” *Internet Interventions*, vol. 25, p. 100422, 2021.
- [11] S. Ghosh, A. Ekbal, and P. Bhattacharyya, “A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes,” *Cognitive Computation*, pp. 1–20, 2022.
- [12] S. Dhelim, L. Chen, H. Ning, and C. Nugent, “Artificial intelligence for suicide assessment using audiovisual cues: A review,” *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5591–5618, 2023.
- [13] E. Moore II, M. A. Clements, J. W. Peifer, and L. Weisser, “Critical analysis of the impact of glottal features in the classification of clinical depression in speech,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 96–107, 2007.
- [14] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, “AVEC 2016: Depression, mood, and emotion recognition workshop and challenge,” in *Proc. ACM-MM*, Amsterdam, 2016.
- [15] W. Wu, C. Zhang, and P. Woodland, “Self-supervised representations in speech-based depression detection,” in *Proc. ICASSP*, Rhodes Island, 2023.
- [16] S. Luz, F. Haider, S. de la Fuente, D. Fromm, and B. MacWhinney, “Alzheimer’s dementia recognition through spontaneous speech: The ADReSS challenge,” in *Proc. Interspeech*, Shanghai, 2020.
- [17] Y. Li, C. Lai, D. Lala, K. Inoue, and T. Kawahara, “Alzheimer’s dementia detection through spontaneous dialogue with proactive robotic listeners,” in *HRI*, 2022, pp. 875–879.
- [18] Z. Cui, W. Wu, W.-Q. Zhang, J. Wu, and C. Zhang, “Transferring speech-generic and depression-specific knowledge for Alzheimer’s disease detection,” in *Proc. ASRU*, Taipei, 2023.
- [19] S. Scherer, J. Pestian, and L.-P. Morency, “Investigating the speech characteristics of suicidal adolescents,” in *Proc. ICASSP*, Vancouver, 2013.
- [20] Z. Cui, C. Lei, W. Wu, Y. Duan, D. Qu, J. Wu, R. Chen, and C. Zhang, “Spontaneous speech-based suicide risk detection using whisper and large language models,” in *Proc. Interspeech*, Kos Island, 2024.
- [21] 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 Mining*, vol. 14, no. 1, pp. 1–17, 2021.
- [22] B. Stasak, J. Epps, H. T. Schatten, I. W. Miller, E. M. Provost, and M. F. Arme, “Read speech voice quality and disfluency in individuals with recent suicidal ideation or suicide attempt,” *Speech Communication*, vol. 132, pp. 10–20, 2021.
- [23] L. Baird, N. Evans, and S. J. Greenhill, “Blowing in the wind: Using ‘north wind and the sun’ texts to sample phoneme inventories,” *Journal of the International Phonetic Association*, vol. 52, no. 3, pp. 453–494, 2022.
- [24] M. I. Conley, D. V. Dellarco, E. Rubien-Thomas, A. O. Cohen, A. Cervera, N. Tottenham, and B. Casey, “The racially diverse affective expression (radiate) face stimulus set,” *Psychiatry research*, vol. 270, pp. 1059–1067, 2018.
- [25] D. V. Sheehan, Y. Lecrubier, K. H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, G. C. Dunbar *et al.*, “The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10,” *Journal of clinical psychiatry*, vol. 59, no. 20, pp. 22–33, 1998.
- [26] D. V. Sheehan, K. H. Sheehan, R. D. Shytle, J. Janavs, Y. Bannon, J. E. Rogers, K. M. Milo, S. L. Stock, and B. Wilkinson, “Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID),” *The Journal of Clinical Psychiatry*, vol. 71, no. 3, p. 17393, 2010.
- [27] Y.-X. Liu, J. Liu, and Y.-F. Wang, “Reliability and validity of Chinese version of the Mini International Neuropsychiatric Interview for Children and Adolescents (child version),” *Chinese Mental Health Journal*, 2011.
- [28] S. McManimen and M. M. Wong, “Prospective investigation of the interaction between social problems and neuropsychological characteristics on the development of suicide ideation,” *Suicide and Life-Threatening Behavior*, vol. 50, no. 2, pp. 545–557, 2020.
- [29] S. Tayebi Arasteh, T. Arias-Vergara, P. A. Pérez-Toro, T. Weise, K. Packhäuser, M. Schuster, E. Noeth, A. Maier, and S. H. Yang, “Addressing challenges in speaker anonymization to maintain utility while ensuring privacy of pathological speech,” *Communications Medicine*, vol. 4, no. 1, p. 182, 2024.
- [30] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, “Wav2vec 2.0: A framework for self-supervised learning of speech representations,” in *Proc. NeurIPS*, Vancouver, 2020.
- [31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional Transformers for language understanding,” in *Proc. NAACL-HLT*, Minneapolis, 2019.