

DAIC-WOZ Dataset Summary

Distress Analysis Interview Corpus - Wizard of Oz

Maintainer: USC Institute for Creative Technologies (ICT)

URL: <https://dcapswoz.ict.usc.edu/>

Access: Research use only (requires data use agreement)

Overview

The DAIC-WOZ (Distress Analysis Interview Corpus - Wizard of Oz) is a **clinical interview dataset** designed to support diagnosis of psychological distress conditions including:

- Anxiety
- Depression
- Post-Traumatic Stress Disorder (PTSD)

Key Characteristics

Attribute	Value
Participants	189 sessions
Duration	7-33 minutes (avg: 16 minutes)
Language	English (American)
Population	US Army veterans and general population
Ground Truth	PHQ-8 questionnaire scores
Interviewer	"Ellie" - animated virtual interviewer
Control Method	Wizard-of-Oz (human interviewer in another room)

Data Collection

The SimSensei Kiosk

Interviews were conducted using the **SimSensei Kiosk** - a virtual human interviewer for healthcare decision support (DeVault et al., 2014).

- **Virtual Agent:** "Ellie" - animated avatar
- **Interview Style:** Semi-structured clinical interview
- **Questions:** Open-ended about life, relationships, stress
- **Wizard-of-Oz:** Human controls Ellie from another room
- **Later versions:** Fully autonomous AI agent (Extended DAIC)

Collection Goals

Part of larger effort to create AI that:

1. Interviews people naturally
2. Identifies verbal and nonverbal indicators of mental illness

- 3. Supports clinical decision-making

Available Modalities

1. Audio

- Full participant audio recordings
- High-quality speech capture

2. Video

- Face video recordings
- Used for facial expression analysis

3. Transcripts

- Full transcription of interactions
- Time-aligned to audio/video

4. Extracted Features

- **Facial Features:** Facial Action Units (FAUs)
- **Acoustic Features:** Pre-extracted using standard toolkits
- **Linguistic Features:** Text-based features

5. Questionnaire Responses

- PHQ-8 (Primary Outcome)
- Additional psychological assessments

PHQ-8 Depression Scoring

The **PHQ-8** (Patient Health Questionnaire-8) is the primary label:

Score Range	Severity
0-4	None/minimal
5-9	Mild
10-14	Moderate
15-19	Moderately severe
20-24	Severe

Clinical Cutoff: PHQ-8 ≥ 10 typically indicates clinically relevant depression

Dataset Splits

AVEC Challenge Partitions

Set	Sessions	Purpose
-----	----------	---------

Training	~107	Model development
Development	~35	Hyperparameter tuning
Test	~47	Final evaluation

Note: Exact splits vary by challenge year (AVEC 2016, 2017, 2019)

AVEC Challenge History

DAIC-WOZ has been used in multiple AVEC (Audio/Visual Emotion Challenge) competitions:

AVEC 2016

- **Task:** Depression severity prediction (PHQ-8)
- **Metric:** Mean Absolute Error (MAE), RMSE

AVEC 2017

- **Extended tasks:** Added PTSD prediction
- **Additional:** GAD-7 for anxiety

AVEC 2019 (Extended DAIC)

- **Key Change:** Test set uses fully autonomous AI interviewer
- **Research Question:** How does AI-only interviewer affect detection?
- **Metric:** Concordance Correlation Coefficient (CCC)

Data Access

How to Obtain

1. Visit: <https://dcapswoz.ict.usc.edu/>
2. Sign data use agreement (for research purposes only)
3. Download individual session ZIP files

File Structure (per session)

```

XXX_P.zip/
|—— XXX_AUDIO.wav      # Participant audio
|—— XXX_TRANSCRIPT.csv  # Time-aligned transcript
|—— XXX_FORMANT.csv     # Formant features
|—— XXX_COVAREP.csv     # COVAREP acoustic features
|—— XXX_FAC.csv        # Facial action coding
|—— XXX_CLNF.csv       # CLnF facial features
|—— XXX_OpenFace.csv   # OpenFace features
|—— metadata/          # Session metadata, labels

```

Total Size

- Individual sessions: 187MB - 957MB each
- Full corpus: ~80GB+

Baseline Results

AVEC Challenge Baselines

Challenge	Modality	Model	MAE (PHQ-8)
AVEC 2016	Audio	SVM	6.74
AVEC 2016	Video	RF	6.12
AVEC 2017	Audio	RF	5.72
AVEC 2019	Multi	DepAudioNet	5.29

State-of-the-Art Results (Literature)

Paper	Year	Model	Performance
Fan et al.	2019	LASSO	MAE: 5.31
Srimadhur & Lalitha	2020	CNN-BiLSTM	MAE: 4.28
Multimodal	Various	Transformer	MAE: ~4.0

Comparison: DAIC-WOZ vs ANDROIDS

Aspect	DAIC-WOZ	ANDROIDS
Language	English	Italian
Labels	PHQ-8 (self-report)	Psychiatric diagnosis
Participants	189 sessions	118 participants
Speech Type	Spontaneous only	Read + Spontaneous
Interview	Virtual agent	Human interviewer
Availability	Restricted	Publicly available
Challenge	AVEC 2016-2019	INTERSPEECH 2023

Implications

- DAIC-WOZ: Larger, self-report labels, controlled setting
 - ANDROIDS: Psychiatric ground truth, multi-task, public access
-

Strengths

1. **Large-scale:** 189 sessions, substantial data volume
2. **Multimodal:** Audio, video, text, extracted features
3. **Well-benchmarked:** Standard AVEC challenge comparisons
4. **English:** Largest English depression speech dataset
5. **Real-world:** Veterans with actual psychological conditions
6. **Pre-extracted features:** Ready for ML experimentation

Limitations

1. **Self-report labels:** PHQ-8 vs clinical diagnosis
2. **Data access:** Requires agreement, not fully public

3. **US-centric:** May not generalize to other populations
4. **Virtual interviewer:** Affects naturalness of speech
5. **No read speech:** Only spontaneous/conversational

Key References

Primary Dataset Paper

```
@inproceedings{gratch2014distress,  
  title={The distress analysis interview corpus of human and computer interviews},  
  author={Gratch, Jonathan and Artstein, Ron and Lucas, Gale M and Stratou, Giota and Scherer, Stefan  
and Nazarian, Angela and Wood, Rachel and Boberg, Jill and DeVault, David and Marsella, Stacy and  
others},  
  booktitle={Proceedings of LREC},  
  pages={3123--3128},  
  year={2014}  
}
```

SimSensei Kiosk

```
@inproceedings{devault2014simsensei,  
  title={SimSensei kiosk: A virtual human interviewer for healthcare decision support},  
  author={DeVault, David and Artstein, Ron and Benn, Grace and Dey, Teresa and Fast, Ed and Gainer,  
Alesia and Georgila, Kallirroi and Gratch, Jon and Hartholt, Arno and Lhommet, Margaux and others},  
  booktitle={Proceedings of AAMAS},  
  year={2014}  
}
```

AVEC 2019 Challenge

```
@inproceedings{ringeval2019avec,  
  title={AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural  
affect recognition},  
  author={Ringeval, Fabien and Schuller, Björn and Valstar, Michel and Cummins, Nicholas and Cowie,  
Roddy and Tavabi, Leili and Schmitt, Maximilian and Alisamir, Sina and Amiriparian, Shahin and Messner,  
Eva-Maria and others},  
  booktitle={Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop},  
  pages={3--12},  
  year={2019}  
}
```

Implications for Nile's Dissertation

Why DAIC-WOZ is Valuable:

1. **Gold standard:** Primary benchmark for depression detection
2. **English:** Relevant to UK/NHS context
3. **Well-documented:** Extensive research history
4. **Feature-ready:** Pre-extracted features available

5. **Comparable:** Standard AVEC metrics allow comparison

Potential Research Directions:

1. **Apply Glasgow methods:** Test Tao's MLA/CDMA on DAIC-WOZ
2. **Cross-corpus validation:** Train ANDROIDS → Test DAIC-WOZ
3. **Feature analysis:** Which acoustic features predict PHQ-8?
4. **Interpretability:** Explain model decisions
5. **NHS applicability:** Can models detect UK depression patterns?

Technical Considerations:

1. **Data agreement required:** Allow time for access
2. **Large downloads:** ~80GB total corpus
3. **PHQ-8 vs clinical:** Consider label reliability
4. **Spontaneous only:** No read speech comparison possible

Related Datasets

Dataset	Language	N	Labels	Public
DAIC-WOZ	English	189	PHQ-8	Restricted
ANDROIDS	Italian	118	Psychiatric	Yes
SEWA	Multi	1900+	Emotion	Restricted
CMU-MOSEI	English	1000+	Emotion	Yes
EATD-Corpus	Chinese	162	SDS	Yes

Document created: 2026-02-01

For Nile's dissertation: "Identifying Depression Through Speech"