

ANDROIDS Corpus - Data Documentation

Overview

Location: ~/Documents/research-workspace/dissertation/data/Androids-Corpus/

Total Size: ~3.69 GB

Audio Format: WAV, 16-bit, mono, 44100 Hz

Dataset Statistics

Task	Healthy Controls (HC)	Patients (PT)	Total
Reading Task	54	58	112
Interview Task	52	64	116

Note: Not all participants completed both tasks. 110 participants have both read and spontaneous speech.

Directory Structure

```
Androids-Corpus/
├── Reading-Task/
│   └── audio/
│       ├── HC/      # 54 healthy control recordings
│       └── PT/      # 58 patient recordings
└── Interview-Task/
    ├── audio/      # Full interview recordings
    │   ├── HC/      # 52 healthy controls
    │   └── PT/      # 64 patients
    └── audio_clip/ # Individual turns (874 clips)
        └── [speaker_id]/ # 116 speaker directories
    ├── Androids.conf # OpenSMILE configuration
    ├── fold-lists.csv # Cross-validation folds
    └── interview_timedata.csv # Turn segmentation
```

File Naming Convention

Format: nn_XGmm_t.wav

Component	Meaning	Values
nn	Speaker ID (unique within group)	01-64
X	Condition	P = Patient, C = Control
G	Gender	M = Male, F = Female
mm	Age	Two digits
t	Education level	1-4 (1=primary, 4=university)

Examples:

- 01_CF56_1.wav → Control, Female, 56 years old, primary education
- 03_PM65_2.wav → Patient, Male, 65 years old, secondary education

Unique identifier: nn_X (e.g., 01_C or 01_P)

Audio Specifications

- **Format:** WAV (RIFF, little-endian)
 - **Encoding:** Microsoft PCM, 16-bit
 - **Channels:** Mono
 - **Sample Rate:** 44100 Hz
-

Tasks

Reading Task

Participants read "The Wind of the North and the Sun" (Aesop's fable) in Italian.

- Controlled text → standardised phonetic content
- Removes semantic/content variation
- Duration: ~50 seconds average

Interview Task

Participants answered questions about daily life.

- Spontaneous speech
 - Natural conversation patterns
 - Duration: ~230 seconds average
-

Provided Files

Androids.conf

OpenSMILE configuration file for feature extraction.

- Use with: SMILEExtract -C Androids.conf -I input.wav -O output.csv
- Extracts: RMSE, MFCC 1-12, ZCR, VP, F0 + deltas (32 features)

fold-lists.csv

Pre-defined 5-fold cross-validation splits.

- Speaker-independent (no data leakage)
- Balanced classes per fold

interview_timedata.csv

Turn-by-turn segmentation for interview recordings.

- Start/end times for each speaking turn
 - Useful for fine-grained analysis
-

Labels

Ground Truth: Professional psychiatric diagnosis

- NOT self-report questionnaires
- Binary: Depressed (PT) vs Control (HC)

Depression types in corpus:

- Major depressive disorder (22)
 - Bipolar disorder, depressive phase (15)
 - Reactive depression (8)
 - Endo-reactive depression (7)
 - Anxiety-depressive disorder (5)
 - Persistent depressive disorder (1)
 - Unspecified (6)
-

For This Project

What we'll use:

1. All Reading Task audio (112 files)
2. All Interview Task audio (116 files)
3. OpenSMILE config (Androids.conf)
4. Fold lists for cross-validation

Our analysis:

- Extract features using OpenSMILE
 - Train classifiers (SVM, Random Forest)
 - Analyse feature importance (SHAP)
 - Compare read vs spontaneous speech
-

Documentation created: February 2, 2026