

Glasgow PhD Thesis Summary

"Speech-based Automatic Depression Detection via Biomarkers Identification and AI Approaches"

Author: Fuxiang Tao

Supervisor: Professor Alessandro Vinciarelli

Institution: University of Glasgow, School of Engineering

Date: August 2023 (submitted), 2024 (published)

Pages: 156

DOI: 10.5525/gla.thesis.84055

URL: <https://theses.gla.ac.uk/84055/>

Abstract Summary

Depression affects 300+ million people globally. Traditional diagnosis is time-consuming and depends on clinical experience. This thesis shows TWO ways to benefit from automatic detection:

1. **Identifying speech markers** (duration, pauses, correlation matrices)
 2. **Novel deep learning models** (Multi-local Attention, Cross-Data Multilevel Attention)
-

Thesis Structure

Chapter	Title	Content
1	Introduction	Background, thesis statements, contributions
2	Background	Cognition in depression, speech production, automatic detection overview
3	The Androids Corpus	New publicly available dataset
4	Speech Duration & Silences	Timing-based markers
5	Feature Correlation Matrices	Acoustic feature relationships
6	Multi-local Attention	Novel attention mechanism
7	Cross-Data Multilevel Attention	Combining read + spontaneous speech
8	Conclusions	Summary, limitations, future work

The ANDROIDS Corpus (Chapter 3)

This is the dataset Nile mentioned exploring!

Overview

- **URL:** <https://github.com/androidscorpus/data>
- **118 participants:** 64 depressed, 54 controls
- **Labels:** Given by professional psychiatrists (NOT self-report questionnaires)

- **Language:** Italian
- **Setting:** Mental health centers (in-the-wild, laptop microphone)

Tasks

1. **Reading Task (RT):** Read "The Wind of the North and the Sun" (Aesop fable)
 - Total: 1h 33m 49s
 - Average: 50.3 ± 10.3 seconds
2. **Interview Task (IT):** Answer questions about daily life
 - Total: 7h 24m 22s
 - Average: 229.8 ± 86.6 seconds

Demographics

- **Age:** 47.3 ± 12.2 years (no significant difference between groups)
- **Gender:** ~2.5:1 female:male ratio (consistent with epidemiology)
- **Education:** Balanced between groups
- **110 participants** have BOTH read and spontaneous speech

Depression Types in Corpus

- 22 major depressive disorder
- 15 bipolar disorder (depressive phase)
- 8 reactive depression
- 7 endo-reactive depression
- 5 anxiety-depressive disorder
- 1 persistent depressive disorder
- 6 unspecified

Advantages Over Other Datasets

1. Professional psychiatric diagnosis (not questionnaire scores)
2. In-the-wild recording conditions
3. Both read AND spontaneous speech from same speakers
4. Matched demographics between groups
5. Manual turn segmentation for conversations
6. Reproducible experimental protocols
7. Italian language (underrepresented)

Technical Methods

Feature Extraction (OpenSMILE)

- **Window:** 25ms, **Step:** 10ms
- **16 base features:**
 - Root Mean Square Energy (RMSE)
 - MFCC 1-12 (12 features)
 - Zero Crossing Rate (ZCR)

- Voicing Probability (VP)
- Fundamental Frequency (F0)
- **32 total features** (16 + deltas)

Baseline Approaches

1. SVM Baseline (BL_SVM)

- Average feature vectors over recording
- Linear kernel SVM

2. LSTM Baseline (BL_LSTM)

- Segment into frames (M=128, half overlap)
- LSTM classifier on each frame
- Majority vote aggregation

Baseline Results

Task	SVM Accuracy	LSTM Accuracy
Read	69.7 ± 6.6%	83.4 ± 2.6%
Interview	64.7 ± 6.3%	81.6 ± 1.6%

Comparable to General Practitioners (57.9-73.1%)

Experimental Protocol

- 3-fold cross-validation
- Speaker-independent (no leakage)
- 10 repetitions for LSTM (random init)
- Tesla T4 GPU

Key Contributions by Chapter

Chapter 4: Duration & Silences

- Depressed speakers have different temporal patterns
- Longer pauses, different speech-to-silence ratios
- Statistically significant differences between groups

Chapter 5: Feature Correlation Matrices

- Correlation between acoustic features differs in depression
- "Stability" measure as biomarker
- Improved performance over baseline

Chapter 6: Multi-local Attention (MLA)

- Novel attention mechanism for depression-relevant information
- Improves accuracy AND confidence
- Reduces time needed for detection

Chapter 7: Cross-Data Multilevel Attention (CDMA)

- Combines read AND spontaneous speech
 - Multiple attention mechanisms
 - Captures both task-specific and common depression markers
 - Best results in thesis
-

Supporting Publications

1. **INTERSPEECH 2023**: "The Androids Corpus: A New Publicly Available Benchmark"
 2. **INTERSPEECH 2020**: "Spotting the Traces of Depression in Read Speech"
 3. **ICASSP 2023**: "Multi-Local Attention for Speech-Based Depression Detection"
 4. *Under review*: "The Relationship Between Speech Features Changes When You Get Depressed"
 5. *To submit*: "Cross-Data Multilevel Attention Mechanisms for Depression Detection"
-

Implications for Nile's Dissertation

Why This Thesis is Valuable:

1. **Same university** - Glasgow standards and expectations
2. **Same topic** - Direct comparison possible
3. **ANDROIDS corpus** - Dataset Nile is exploring
4. **Recent work** - 2023/2024, state of the art
5. **Clear methods** - Reproducible protocols
6. **Published papers** - Citable, validated work

Potential Angles for Nile:

1. **Replicate on DAIC-WOZ** - Apply Tao's methods to English dataset
2. **Compare datasets** - ANDROIDS vs DAIC-WOZ cross-validation
3. **Extend CDMA** - Add linguistic features or other modalities
4. **Interpretability** - Which features/attention weights matter most?
5. **UK focus** - Apply to English speakers, NHS context

Key Differences to Consider:

- Tao used Italian data; Nile likely using English (DAIC-WOZ)
 - Tao had psychiatric labels; DAIC-WOZ uses PHQ-8 questionnaire
 - Different evaluation metrics may be needed
-

Contact

Supervisor: Professor Alessandro Vinciarelli (may be worth reaching out if same school)