Background (Draft)

This is a draft of the Background chapter for your dissertation. LaTeX version also available.

This chapter provides the theoretical and technical foundation for the present work.

1. Depression: Clinical Overview

Definition and Prevalence

Major Depressive Disorder (MDD), commonly referred to as depression, is a mental health condition characterised by persistent feelings of sadness, hopelessness, and loss of interest in activities. According to the World Health Organization, depression affects over **280 million people globally**, making it one of the leading causes of disability worldwide.

In the **United Kingdom**, approximately **1 in 6 adults** experiences depression or anxiety in any given week. The economic burden is substantial: depression costs the UK economy an estimated **£105 billion annually** through healthcare costs, lost productivity, and reduced quality of life.

Current Diagnostic Methods

Depression diagnosis traditionally relies on clinical interviews and standardised questionnaires:

Instrument	Type	Scoring
PHQ-9/PHQ-8	Self-report	0-27 scale, ≥10 = clinical depression
Beck Depression Inventory	Self-report	21 items measuring symptoms
Hamilton Rating Scale	Clinician-administered	17-21 items

Limitations of current methods:

1. **Subjectivity:** Inter-rater reliability between clinicians can be as low as 0.28 kappa
2. **Self-report bias:** Patients may underreport due to stigma or lack of insight
3. **Resource requirements:** Requires trained professionals, time-consuming
4. **Access barriers:** Many lack access to mental health professionals

These limitations motivate the search for **objective, scalable biomarkers**—among which speech has emerged as promising.

---

2. Depression and Speech Production

The Cognition-Speech Connection

Speech production is complex, involving planning, motor control, and real-time monitoring. Depression affects multiple cognitive domains that directly influence speech:

- **Psychomotor retardation:** Slowing of processes → reduced speech rate, longer pauses
- **Reduced motivation:** Diminished drive → "flat" or monotonous speech
- **Cognitive load:** Rumination consumes resources → affects fluency
- **Emotional blunting:** Reduced reactivity → decreased pitch variation

Crucially, many changes are involuntary—speakers cannot easily mask them, making speech a potentially robust biomarker.

Observable Speech Changes

Category	Changes in Depression
Temporal	Slower speech rate, longer pauses, more hesitations
Prosodic	Reduced pitch range, lower mean F0, less variability
Voice quality	Increased breathiness, jitter, shimmer
Articulation	Less precise consonants, reduced vowel space

---

3. Acoustic and Prosodic Features

Fundamental Frequency (F0)

F0 corresponds to vocal fold vibration rate, perceived as pitch.

Common F0 features:

- Mean F0: average pitch
- F0 standard deviation: pitch variability
- F0 range: max - min

Depressed speakers typically show lower mean F0 and reduced variability.

Energy and Intensity

Root Mean Square Energy (RMSE) reflects vocal effort:

$$\text{RMSE} = \mathrm{sqrt}(1/N \times \Sigma(x^2))$$

Depressed speakers show reduced overall energy and dynamic range.

Mel-Frequency Cepstral Coefficients (MFCCs)

Spectral features approximating human auditory perception:

1. Compute power spectrum (STFT)
2. Apply mel-scale filterbank
3. Take logarithm
4. Apply Discrete Cosine Transform

Typically 12-13 MFCCs + delta + delta-delta = **39 features**

Temporal Features

- **Speech rate:** Syllables/phonemes per second
- **Pause duration:** Length of silent intervals
- **Speech-to-pause ratio:** Voiced vs unvoiced proportion

Voice Quality Features

- **Jitter:** Cycle-to-cycle F0 variation
- **Shimmer:** Cycle-to-cycle amplitude variation
- **HNR:** Harmonic-to-Noise Ratio (periodic vs aperiodic energy)

Higher jitter/shimmer, lower HNR = less stable phonation, often observed in depression.

Standard Feature Sets

Feature Set	Features	Purpose
eGeMAPS	88	Designed for affective computing
ComParE	6,000+	Comprehensive paralinguistics

This work uses **OpenSMILE** following Tao et al.'s protocol for ANDROIDS.

---

4. Machine Learning Approaches

Traditional ML

Method	Strengths	Use in Depression Detection
SVM	Effective with high-dimensional data	Widely used baseline
Random Forest	Feature importance built-in	Good interpretability
Logistic Regression	Highly interpretable	Coefficient analysis

Deep Learning

Architecture	Strength	Example
LSTM	Temporal dependencies	Tao's baseline
CNN	Spectrogram features	Ma et al., 2016
Attention	Focus on relevant segments	Multi-Local Attention

Evaluation Metrics

- **Accuracy:** Proportion correct
- **F1 Score:** Harmonic mean of precision/recall
- **CCC:** Concordance Correlation Coefficient (regression)
- **AUC:** Area Under ROC Curve

---

5. Related Work

The AVEC Challenges

AVEC 2016, 2017, 2019 used DAIC-WOZ corpus:

- Multimodal approaches outperform unimodal
- Baseline: ~5-6 MAE on PHQ-8
- Deep learning achieves state-of-the-art

The ANDROIDS Corpus

Tao et al. (2023) introduced ANDROIDS addressing key limitations:

Feature	ANDROIDS	DAIC-WOZ
Labels	Psychiatric diagnosis	PHQ-8 self-report
Speech tasks	Read + Spontaneous	Spontaneous only
Language	Italian	English
Access	Public	Restricted

Results: 83.4% accuracy (read), 81.6% (spontaneous)

The Gap: Feature Importance

Most research focuses on **maximising accuracy** ("black box").

Underexplored: Which features drive predictions? This is crucial for:

- Clinical utility (doctors need explanations)
- Scientific understanding (what changes in depression?)
- System design (which features to collect?)

This dissertation addresses this gap.

Summary

- Depression is prevalent with significant diagnostic challenges
- Speech offers a non-invasive, scalable biomarker
- Acoustic features can be formally defined and extracted
- ML approaches range from SVMs to attention networks
- **ANDROIDS provides ideal testbed** with professional labels and dual tasks
- **Gap:** Feature importance analysis remains underexplored

The following chapter describes methodology for analysing feature importance.

Draft version - February 2026
Word count: ~900 (will expand to ~3000 in final)