

Mood State Prediction From Speech Of Varying Acoustic Quality For Individuals With Bipolar Disorder

John Gideon¹, Emily Mower Provost¹, and Melvin McInnis²

Departments of: Computer Science and Engineering¹ and
Psychiatry², University of Michigan



Overview

Bipolar disorder

Pathological mood-state swings of mania and depression
A leading cause of disability – 4% of Americans affected

Current Treatment

Periodic follow-up visits for monitoring
Reactively after manic/depressive episodes

**Costly
Consequences**

Clinical Need

To passively detect & predict mood and health state changes in order to intervene and prevent episodes



National Institute of Mental Health, "Bipolar Disorder In Adults."
Kessler et al., "Lifetime Prevalence And Age-of-onset Distributions Of DSM-IV Disorders In The National Comorbidity Survey Replication."
Angst et al., "Long-term Outcome And Mortality Of Treated Versus Untreated Bipolar And Depressed Patients: A Preliminary Report."



Problem Statement

- **Speech** patterns shown to **reflect mood** in clinic
 - **Controlled environments**
 - Single type of **recording device**
- Real world recordings
 - Variations in **background noise**
 - Variations in **microphone quality**

Speech recorded in the **real world** has **large variations in quality** making a **distributed** mobile health system using speech **infeasible without controlling for these differences.**



UM PRIORI Acoustic Database

- **Participants:** 37 subjects enrolled for 6-12 months
- **Total Data:** 2,400 hours across 30,000 calls
- **Ground Truth:** 780 Recorded weekly phone-based clinical assessments (About 15 minutes each)
 - Structured clinical interview
 - Rated on mania and depression severity
 - Young Mania Rating Scale (**YMRS**)
 - Hamilton Rating Scale for Depression (**HAMD**)
 - 23 assessments transcribed for validating segmentation
 - Only used assessment calls in this analysis

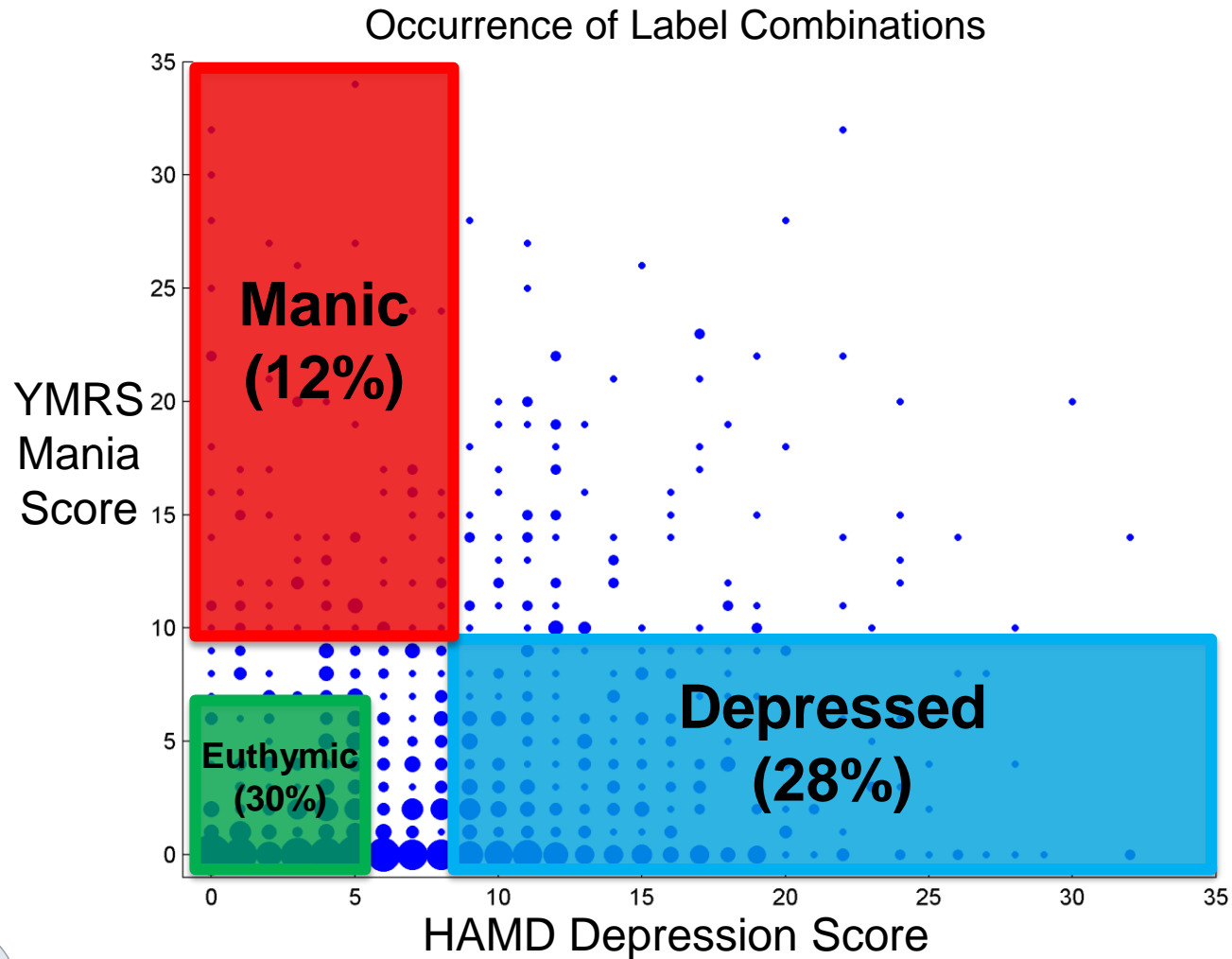
Feelings of guilt?
Insomnia?
Anxiety?
Weight loss?




Hamilton, "Hamilton Depression Scale."
Young et al., "A Rating Scale For Mania: Reliability, Validity And Sensitivity."



Mood Label Assignment



Models of Phones

Samsung Galaxy S3	Samsung Galaxy S5
 A white Samsung Galaxy S3 smartphone is shown vertically. The screen displays a dandelion seed head against a blue sky. The time is 12:03 and the date is Thu, 3 May. The text "swipe screen to unlock" is visible at the bottom of the screen.	 A black Samsung Galaxy S5 smartphone is shown vertically. The screen displays a colorful geometric pattern of triangles in shades of blue, green, yellow, and red. The time is 12:45 and the date is Mon, 24 February. The text "swipe screen to unlock" is visible at the bottom of the screen.
18 Participants	17 Participants
456 Assessments	287 Assessments

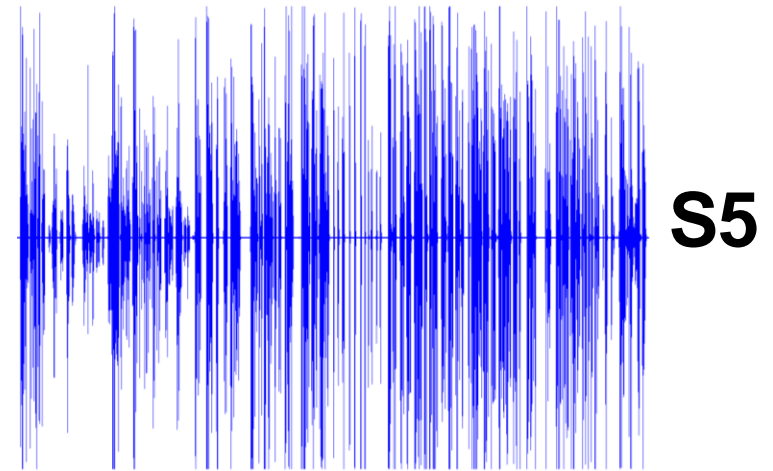
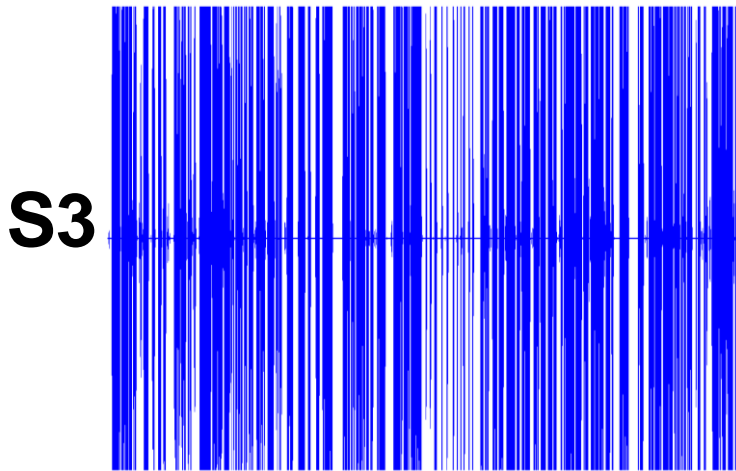
Acoustic Differences Between Models

Galaxy S3 audio
versus S5

Over 100 times
as much **Clipping**

Over 6 times as
loud (**RMS**)

3.9dB drop in
estimated **SNR**



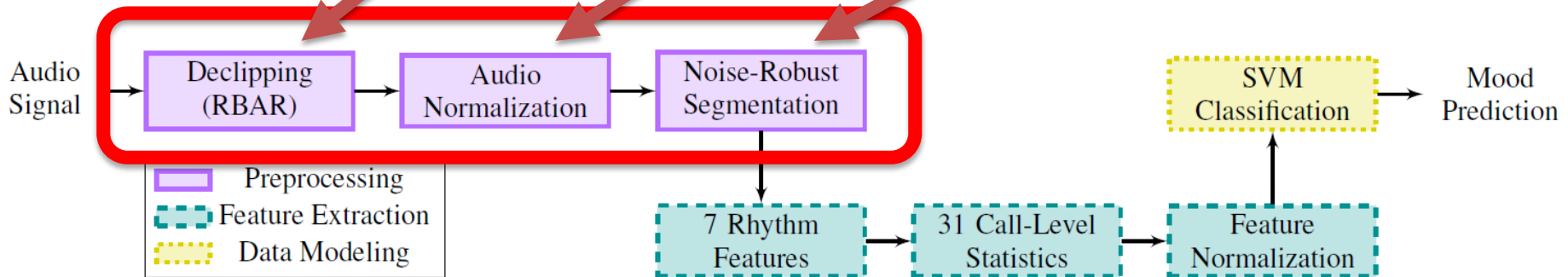
Processing Pipeline – Preprocessing

Galaxy S3 audio
versus S5

Over 100 times
as much **Clipping**

Over 6 times as
loud (**RMS**)

3.9dB drop in
estimated **SNR**

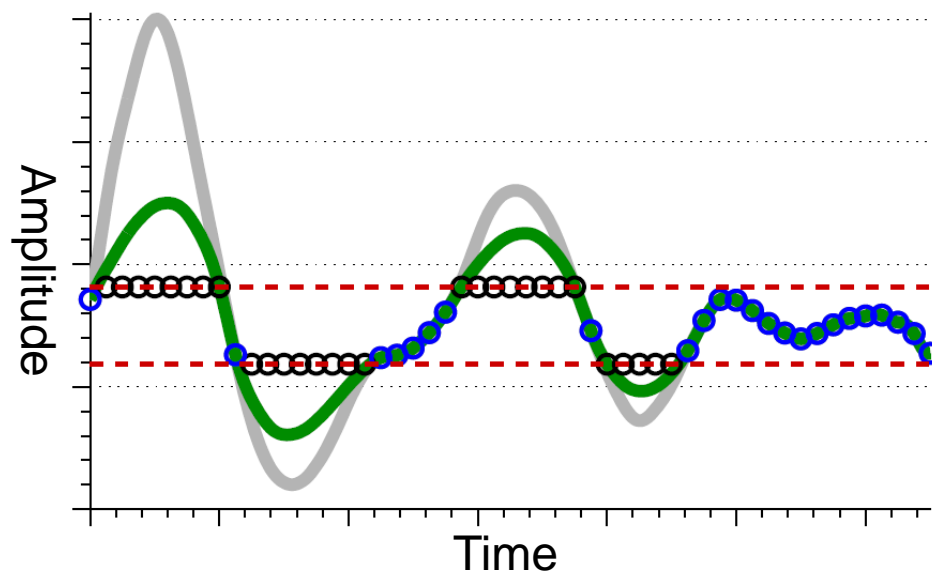


Declipping Method

- **CBAR**

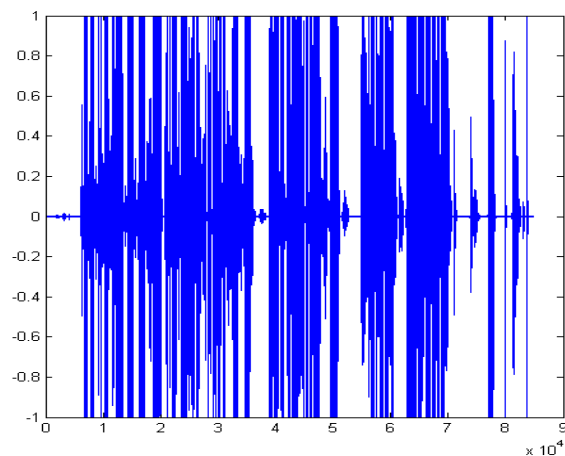
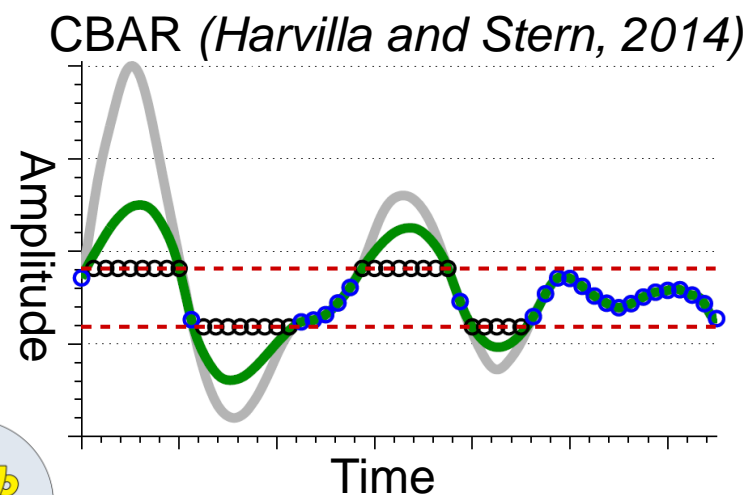
- Extrapolates clipped regions
- Minimizes pointiness (acceleration)

CBAR (*Harvilla and Stern, 2014*)

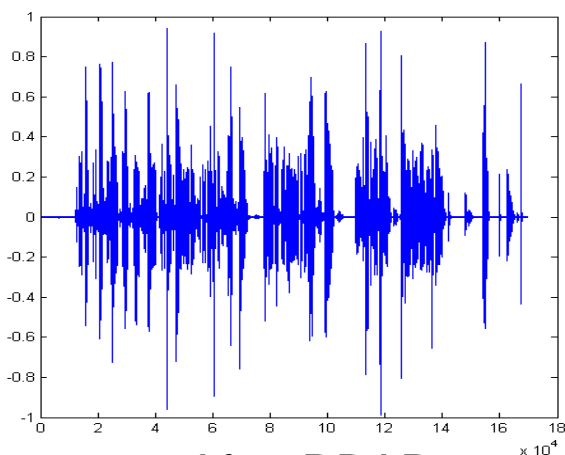


Declicking Method

- **CBAR**
 - Extrapolates clipped regions
 - Minimizes pointiness (acceleration)
- **RBAR**
 - Fast approximation to CBAR
 - Used in preprocessing pipeline

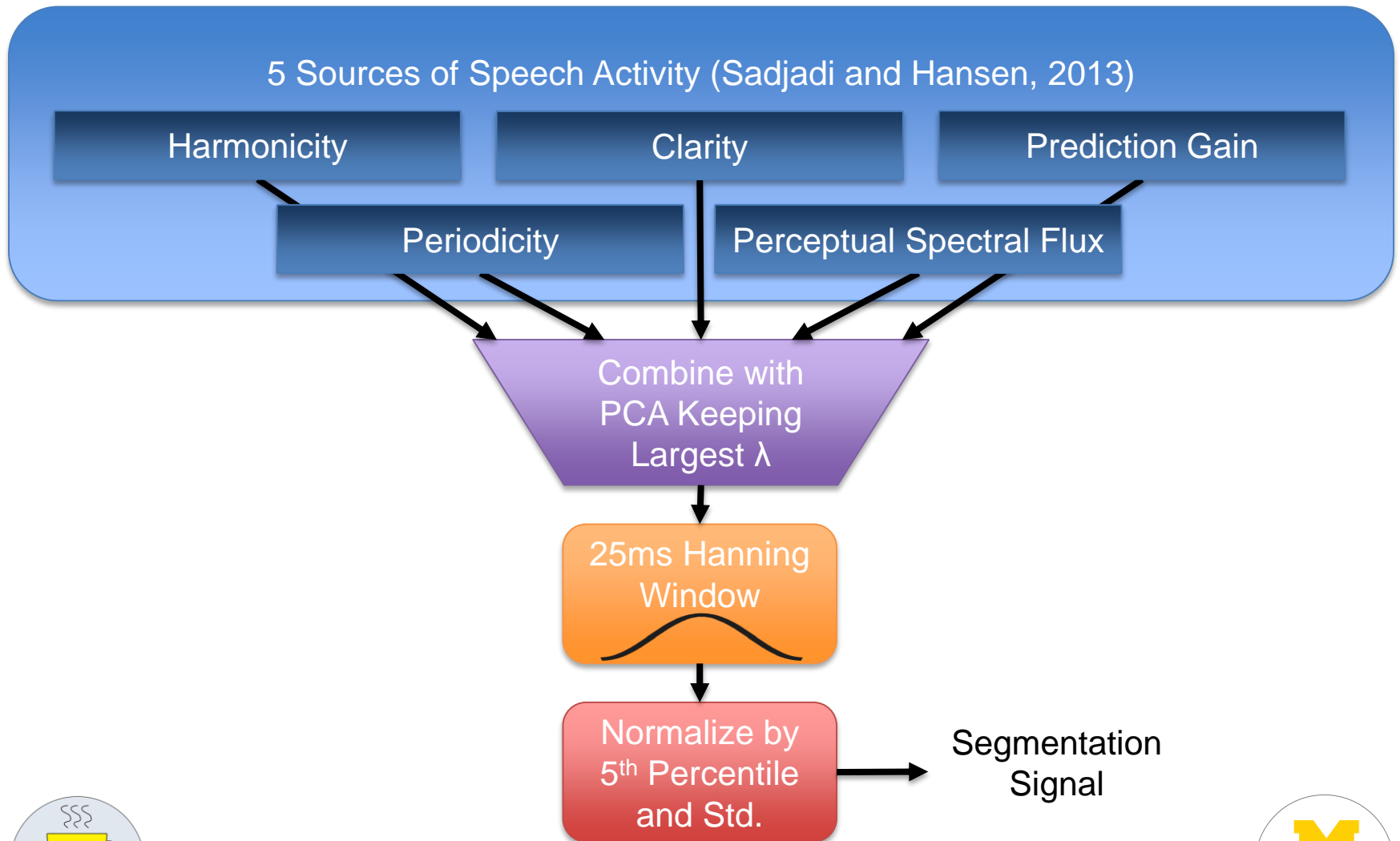


Before



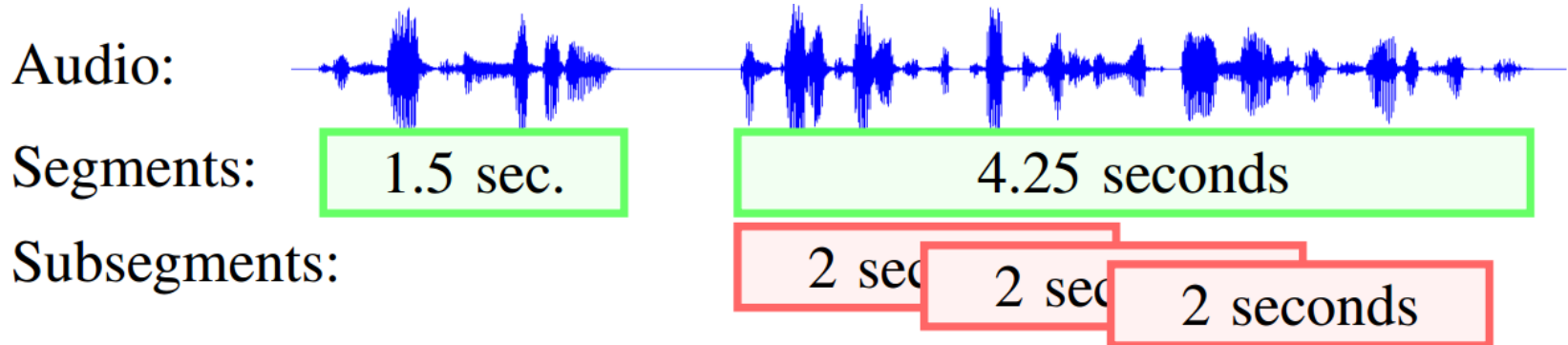
After RBAR

Noise-Robust Segmentation

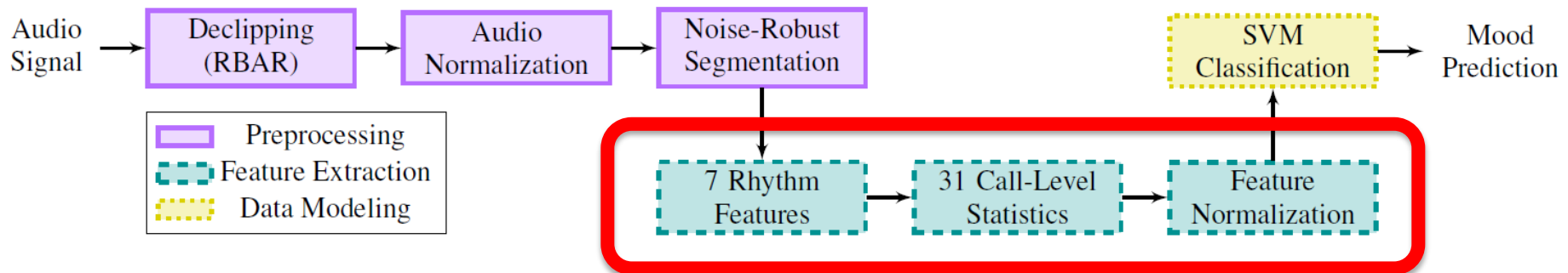


Noise-Robust Segmentation (Cont.)

- Validation used to determine segments
 - Exceeds a **threshold of 1.8**
 - **Minimum silence of 0.7 seconds**
- Only include segments longer than two seconds
 - **Subsegment** into two seconds with one second overlap
 - Necessary for feature extraction

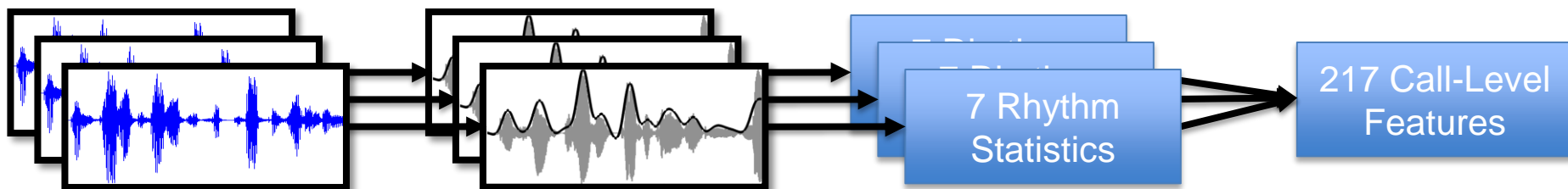


Processing Pipeline – Feature Extraction

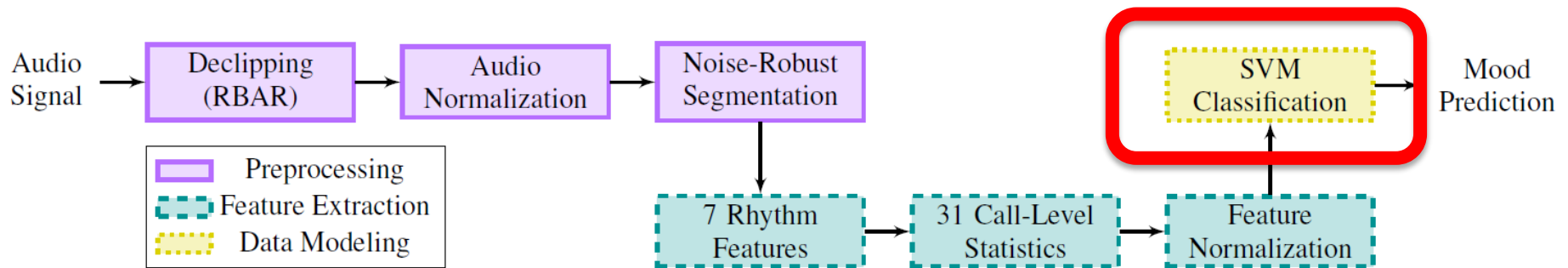


Rhythm Features

- Both mania and depression have rhythm related symptoms
 - **Mania:** Speech is more frequent, quicker, and louder
 - **Depression:** Slowing of speech and difficulty articulating
- Uses constant **two second segments**
 - Extract audio envelope
 - Extract seven statistics of syllable vs supra-syllable rhythm
 - Calculate **31 statistics** over segments for call-level features
- Normalize either **globally** or by **subject**



Processing Pipeline – Data Modeling



Data Partitioning

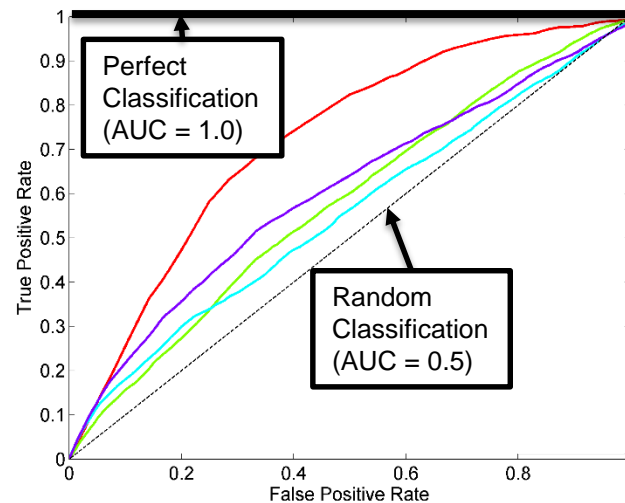
- Binary cases considered
 - Euthymic vs. **manic**
 - Euthymic vs. **depressed**
- Used **participant-independent testing**
- Participants have at least six calls
 - At least two euthymic
 - At least two manic and/or depressed

Model	# Subjects for Mania Test	# Subjects for Depressed Test
S3	12	11
S5	3	7
Both	15	18



Validation, Training, and Testing

- Use **participant-independent validation**
 - Calculate **weighted information gain** and rank features
- Certain experiments use a **Multi-Task SVM**
 - Phone device (S3/S5) is second task
 - Weight kernel function based on device
- Performance measure: **Area Under the Receiver Operating Characteristic Curve (AUC / AUROC)**



Results – Declipping, Normalization, and Multitask

Pipeline Test	Manic AUC	Depressed AUC
Baseline	0.57 ± 0.25	0.64 ± 0.14
Declipped Using RBAR	0.70 ± 0.17*	0.65 ± 0.15
Normalized By Subject	0.67 ± 0.19*	0.75 ± 0.14*
Multi-Task Using Baseline Preprocessing	0.68 ± 0.23*	0.66 ± 0.18
Multi-Task Using Best Preprocessing	0.72 ± 0.20*	0.71 ± 0.15

- **Significantly improved manic performance**
 - S5: Significantly more clipping in manic vs. depressed calls
 - Hypothesis: Individuals speak more loudly in a manic state
- **Normalization by subject** significantly improves both

*Denotes results significantly better than baseline (paired t-test, p=0.05)

Results – No Speech Segmentation

Model	Manic AUC	Depressed AUC
S3	0.52 ± 0.22	0.66 ± 0.17
S5	0.78 ± 0.31	0.62 ± 0.09
Both	0.57 ± 0.25	0.64 ± 0.14

Baseline

Model	Manic AUC	Depressed AUC
S3	0.73 ± 0.22	0.74 ± 0.10
S5	0.79 ± 0.37	0.80 ± 0.21
Both	$0.74 \pm 0.24^*$	$0.77 \pm 0.15^*$

No Speech Segmentation

- **Remove speech segmentation**

- Divide all audio into two second segments with one second overlap
- Silence is included in features

- Accuracy significantly improves

- Hypothesis: Rhythm features **indirectly capturing information** about the assessment interview
- Requirement: **Accurate segmentation to avoid misleading results**

*Denotes results significantly better than baseline (paired t-test, $p=0.05$)



Conclusion

- Results demonstrate ability to counter variations in recording device quality
 - Differences include **clipping, loudness, and noise**
 - Combination of **preprocessing, feature extraction, and data modeling**
- **Significantly better than baseline**
 - Manic: $0.57 \pm 0.25 \rightarrow 0.72 \pm 0.20$
 - Depressed: $0.64 \pm 0.14 \rightarrow 0.75 \pm 0.14$
- **No comprehensive solution**
- Techniques could also be used to increase **subject comparability** when performing analysis on **personal calls**



Thank you for listening!

Questions?

