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

John Gideon[1], Emily Mower Provost[1], and Melvin McInnis[2]

Departments of: Computer Science and Engineering[1] and Psychiatry[2], University of Michigan

CHAI Lab
University of Michigan

COMPUTER SCIENCE
& ENGINEERING
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.**

CHAI Lab
University of Michigan

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

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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?

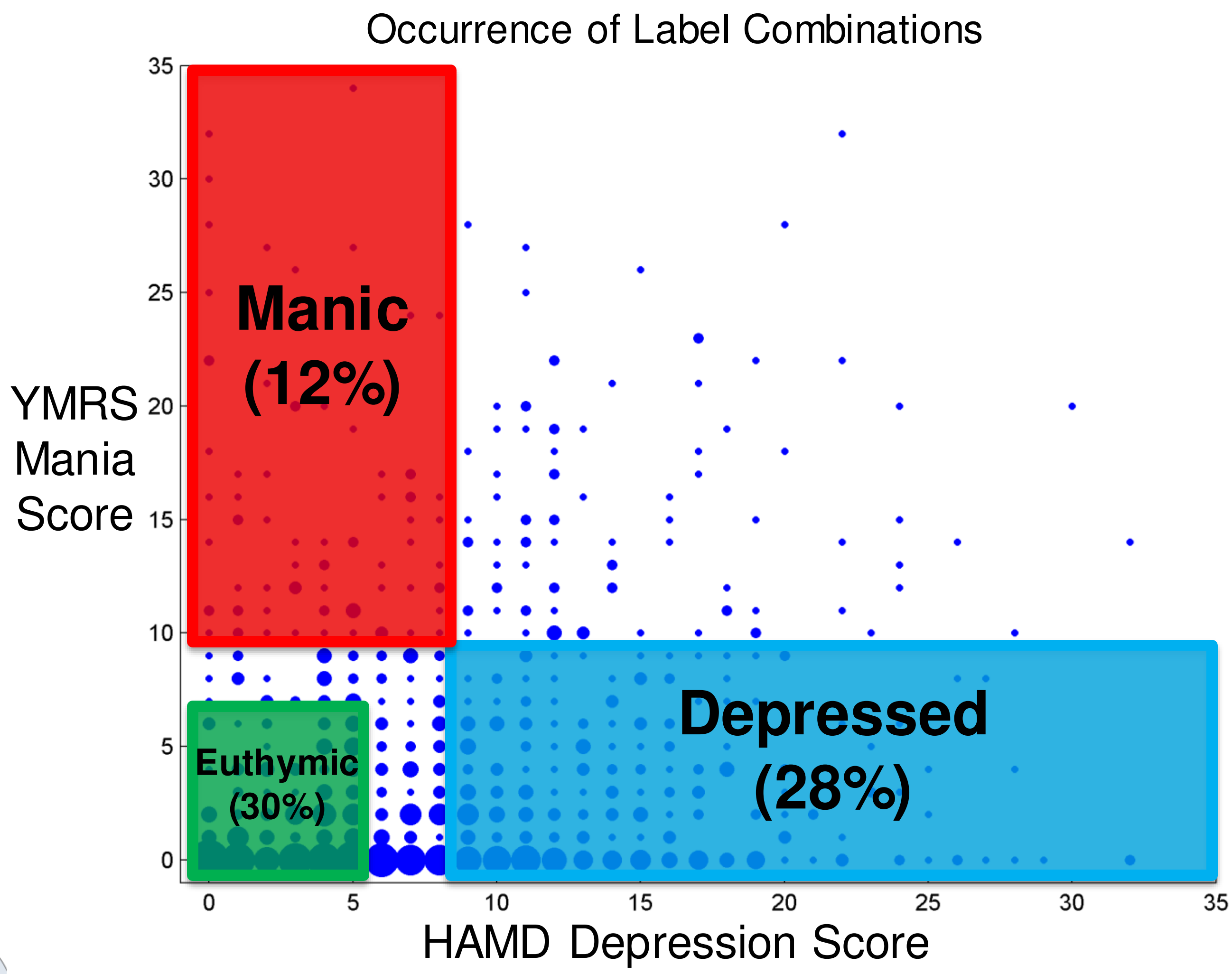
Assessment Call Audio → ? → Assessment Mood

CHAI Lab
University of Michigan

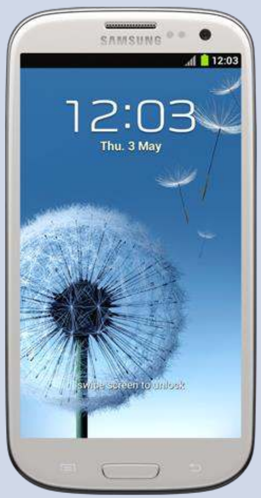
Hamilton, "Hamilton Depression Scale."
Young et al., "A Rating Scale For Mania: Reliability, Validity And Sensitivity."

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Mood Label Assignment



Models of Phones

Samsung Galaxy S3	Samsung Galaxy S5
	
18 Participants	17 Participants
456 Assessments	287 Assessments

Images: Samsung.com

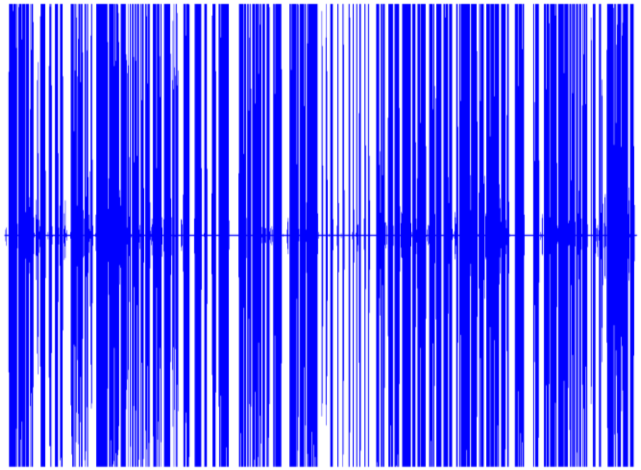
CHAI Lab
University of Michigan

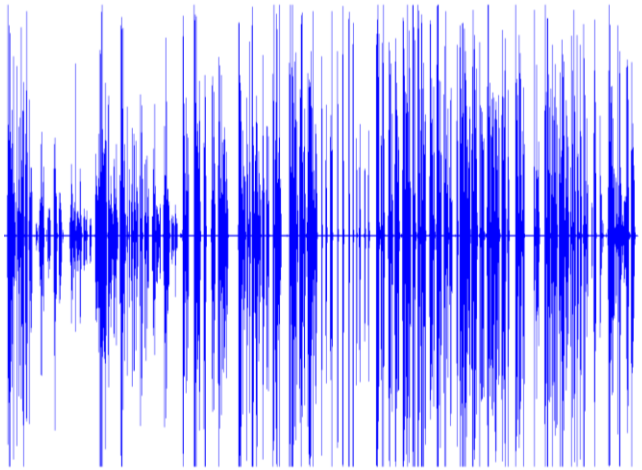
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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**

S3

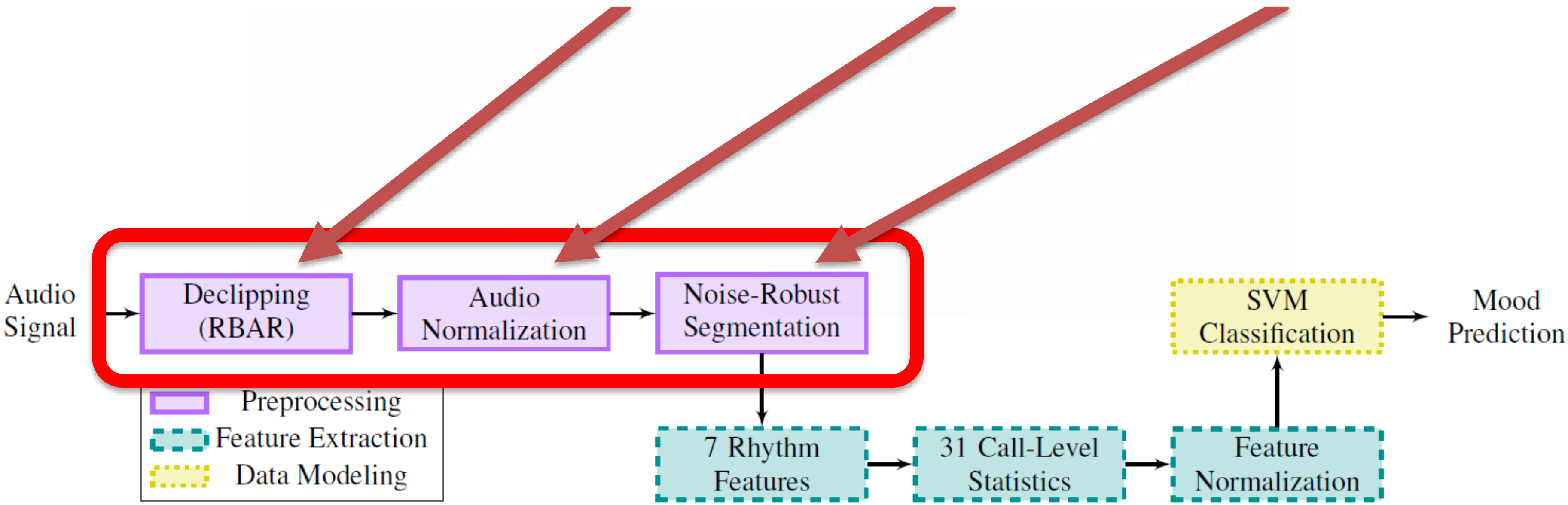




S5

Processing Pipeline – Preprocessing

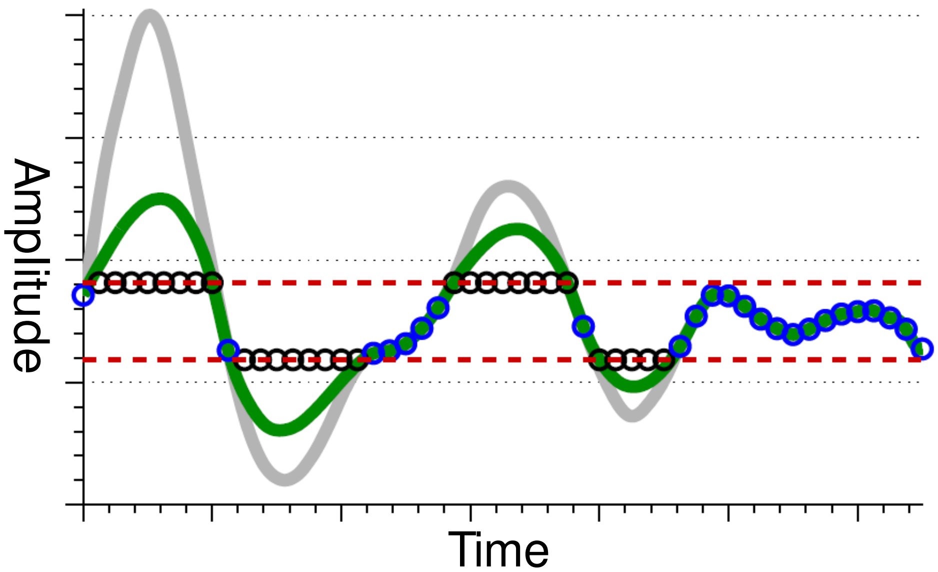
Galaxy S3 audio versus S5	Over 100 times as much **Clipping**	Over 6 times as loud (**RMS**)	3.9dB drop in estimated **SNR**



Declipping Method

- **CBAR**
  - Extrapolates clipped regions
  - Minimizes pointiness (acceleration)

CBAR *(Harvilla and Stern, 2014)*

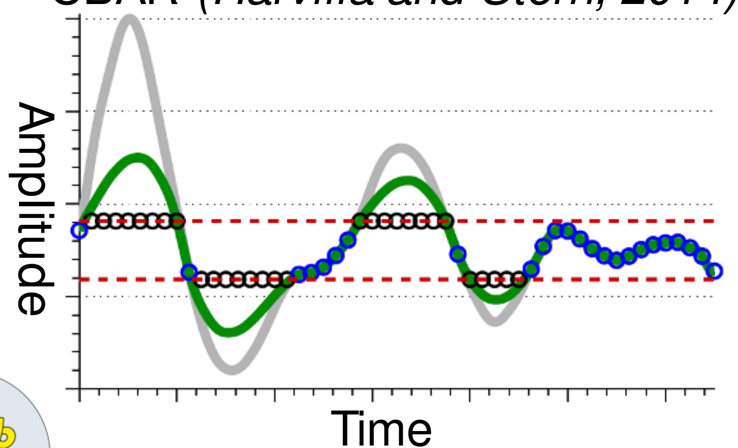


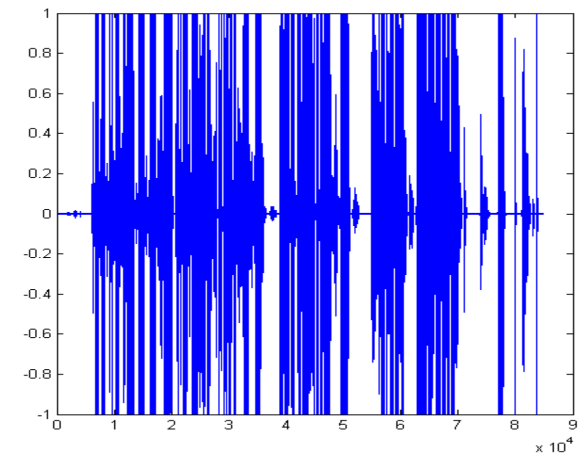
Harvilla and Stern. "Least Squares Signal Declipping For Robust Speech Recognition."
Harvilla and Stern. "Efficient Audio Declipping Using Regularized Least Squares."

Declipping Method

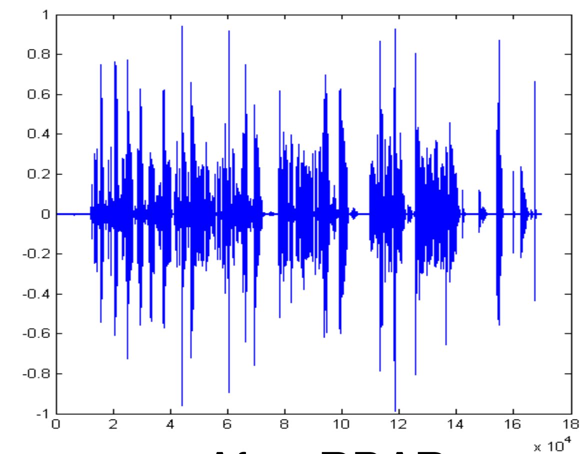
- **CBAR**
  - Extrapolates clipped regions
  - Minimizes pointiness (acceleration)
- **RBAR**
  - Fast approximation to CBAR
  - Used in preprocessing pipeline

CBAR *(Harvilla and Stern, 2014)*





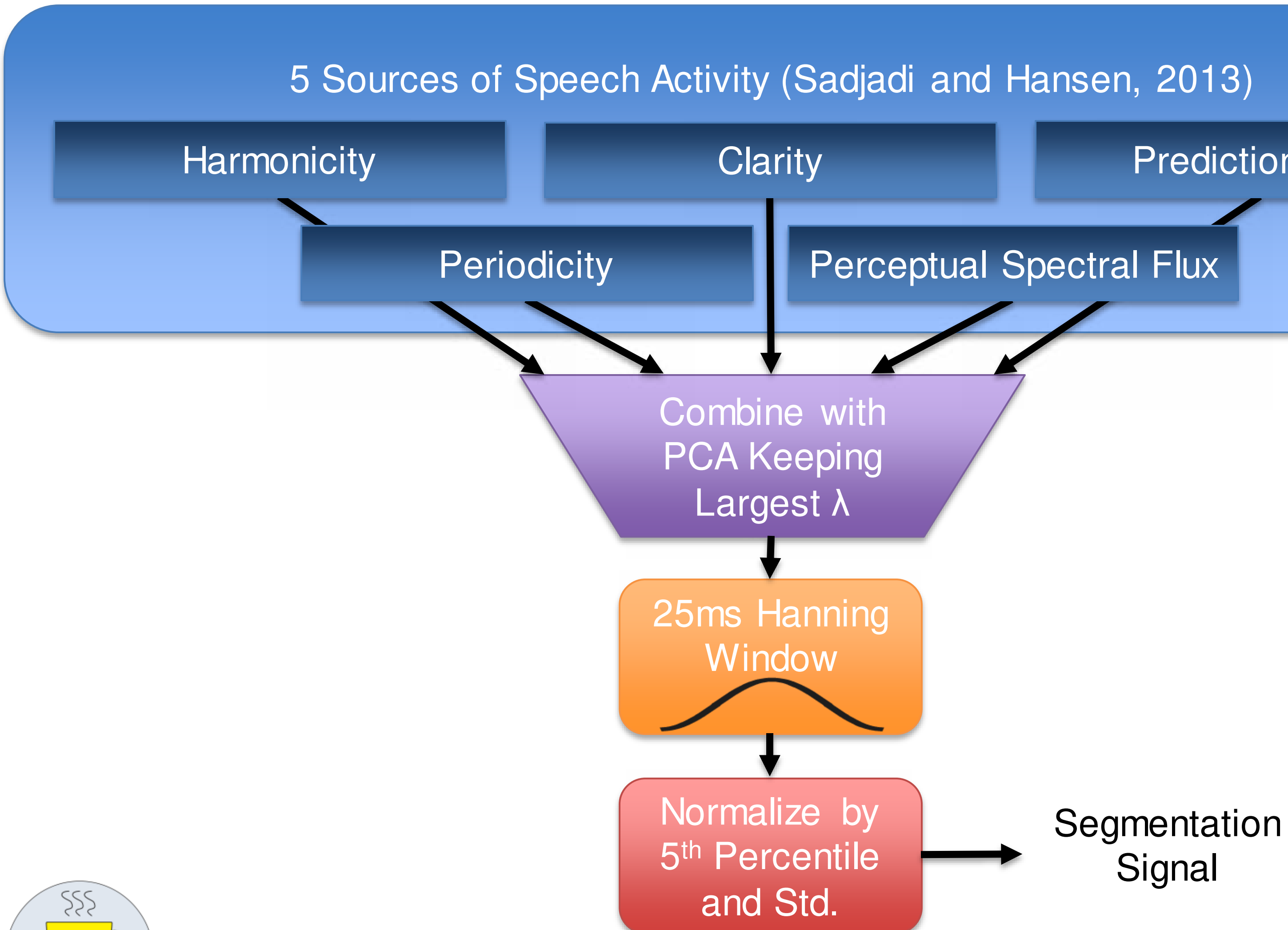
Before



After RBAR

Harvilla and Stern. "Least Squares Signal Declipping For Robust Speech Recognition."
Harvilla and Stern. "Efficient Audio Declipping Using Regularized Least Squares."

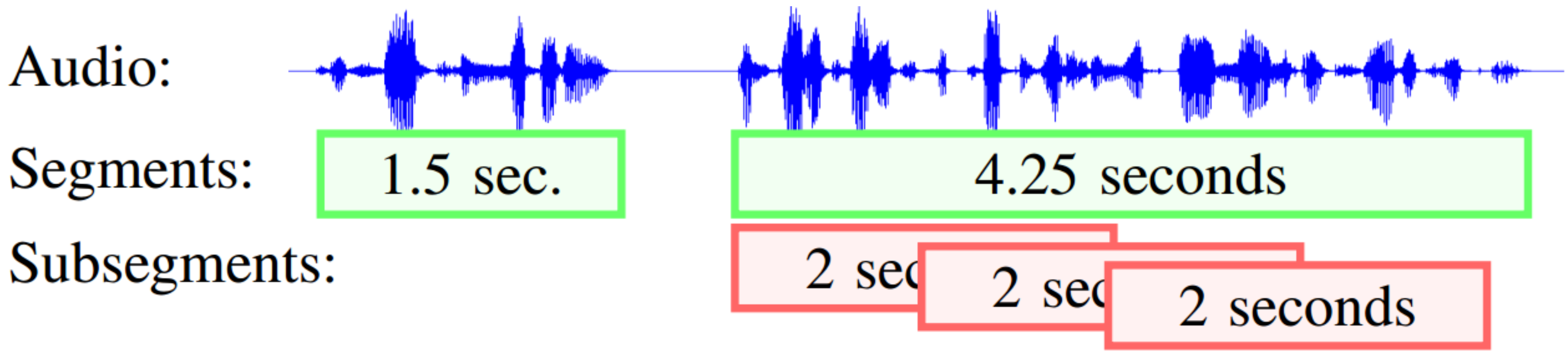
Noise-Robust Segmentation



Sadjadi and Hansen. "Unsupervised Speech Activity Detection Using Voicing Measures And Perceptual Spectral Flux."

Noise-Robust Segmentation (Cont.)

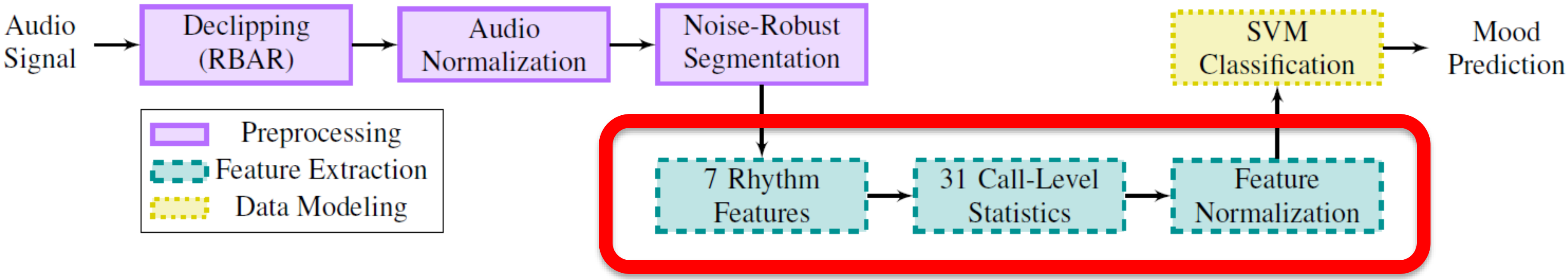
- Validation used to determine segments
  - Exceeds a **threshold of 1.8**
  - **Minimum silence of 0.7 seconds**
- Only include segments longer than two seconds
  - **Subsegment** into two seconds with one second overlap
  - Necessary for feature extraction



CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Processing Pipeline – Feature Extraction

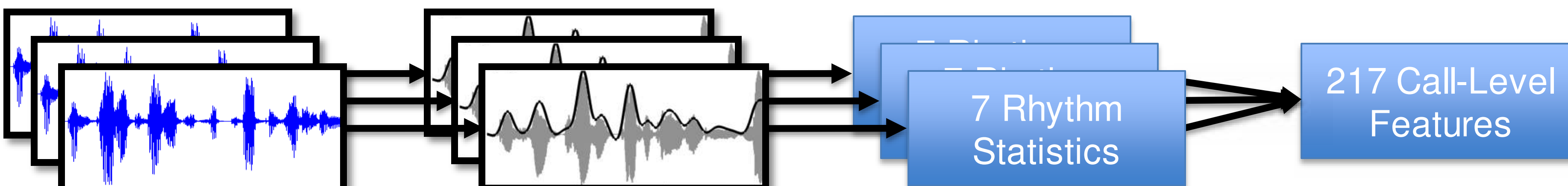


CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

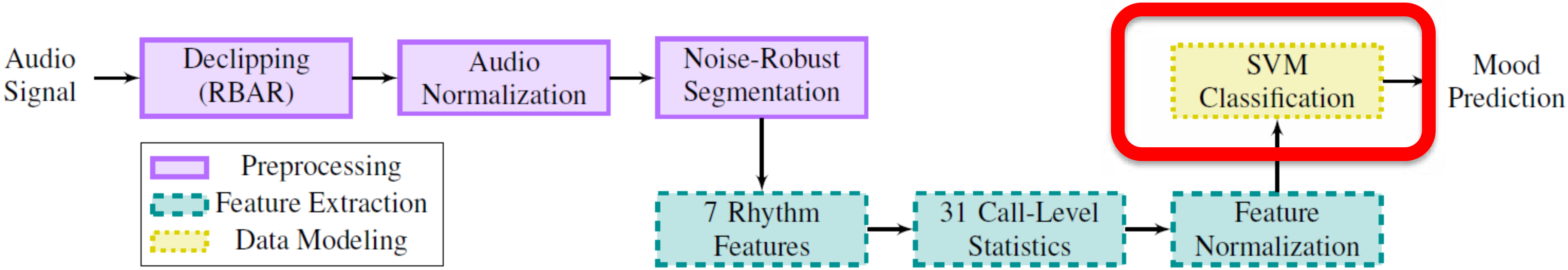
Rhythm Features

- 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**



Tilsen and Arvaniti. "Speech Rhythm Analysis With Decomposition Of The Amplitude Envelope: Characterizing Rhythmic Patterns Within And Across Languages."

Processing Pipeline – Data Modeling



CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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

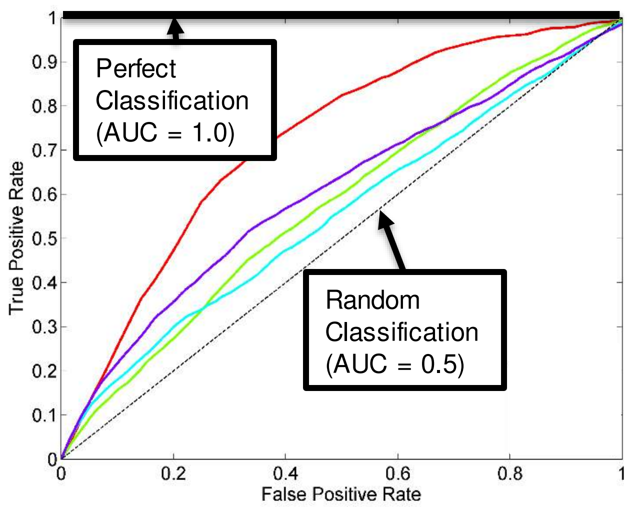
Model	# Subjects for Mania Test	# Subjects for Depressed Test
S3	12	11
S5	3	7
Both	**15**	**18**

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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)**



CHAI Lab University of Michigan

COMPUTER SCIENCE & ENGINEERING UNIVERSITY OF MICHIGAN

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

CHAI Lab
University of Michigan

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

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Results – No Speech Segmentation

Baseline

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

No Speech Segmentation

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 ± 0.24***	**0.77 ± 0.15***

- **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)**

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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 ± 0.25 → **0.72 ± 0.20**
  - Depressed: 0.64 ± 0.14 → **0.75 ± 0.14**
- **No comprehensive solution**
- Techniques could also be used to increase **subject comparability** when performing analysis on **personal calls**

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Thank you for listening!

Questions?

CHAI Lab
University of Michigan

MICHR
Michigan Institute for Clinical & Health Research

NIH National Institutes of Health

The Heinz C. Prechter
Bipolar Research Fund
at the
University of Michigan
Depression Center

COMPUTER SCIENCE
& ENGINEERING
UNIVERSITY OF MICHIGAN

Speech for Mood Monitoring

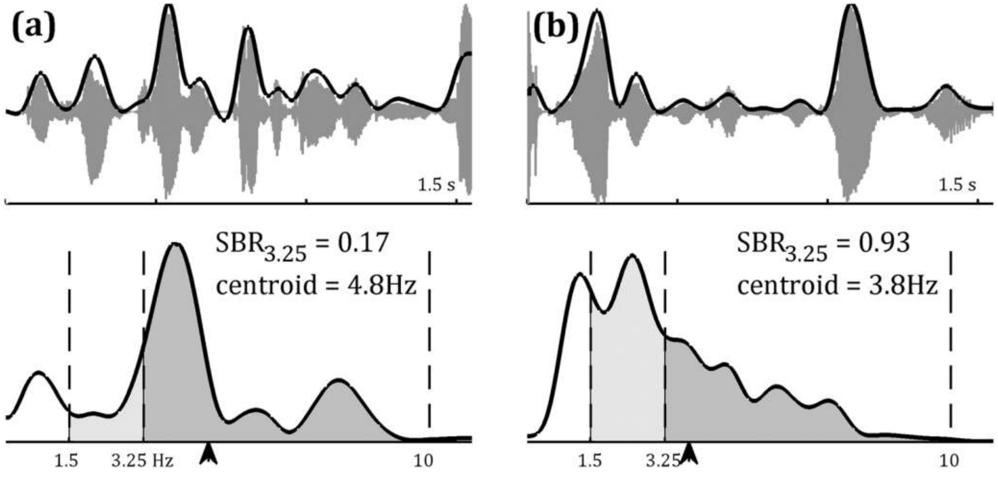
- Computational Analysis of Speech
  - **Emotion Recognition:** Mower 2011, Schuller 2009
  - **Major Depression:** Mundt 2007, Cohn 2009, Trevino 2011, Quatieri 2012, Helfer 2013, Cummins 2013
  - **PTSD:** Sluis 2011, Broek 2011, Tsumatori 2011
  - **Autism:** Hoque 2009, Van Santen 2010, Bone 2012, Chaspari 2013
- Challenges to adoption of remote monitoring
  - Collected in lab or disruptive phone calls
  - Clinical setting: prompted speech, fixed text

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Rhythm Features

- Uses constant **2 second segments**
  - Constant to ensure changes in features due to rhythm, not segment size
  - Provides enough syllables without too much variation
- Perform preprocessing to extract audio envelope (Tilsen, 2013)
- Find power spectra
  - High frequency
    - Syllables
  - Low frequency
    - Supra-syllables



Tilsen and Arvaniti. "Speech Rhythm Analysis With Decomposition Of The Amplitude Envelope: Characterizing Rhythmic Patterns Within And Across Languages."

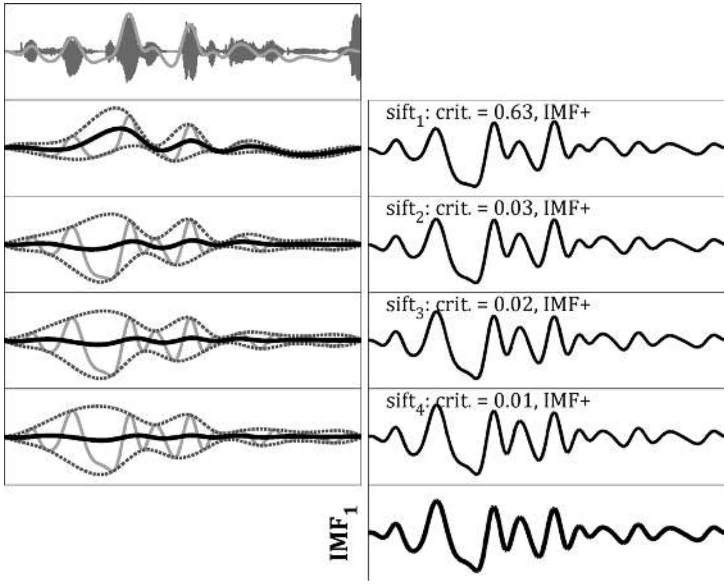
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Rhythm Features (Cont.)

- **Empirical mode decomposition**
  - Extracts the intrinsic mode functions (IMFs)
- Calculate **ratio of power** between IMF_1 and IMF_2
- Determine **instantaneous frequency** over the first two IMFs
  - Time derivative of phase
  - Calculate mean and std.
- Calculate **31 statistics** over segments for call-level features
  - Total Features: 31*7 = **217 total features**
- Normalize either **globally** or by **subject**

EMD



Tilsen and Arvaniti. "Speech Rhythm Analysis With Decomposition Of The Amplitude Envelope: Characterizing Rhythmic Patterns Within And Across Languages."

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Results – Declipping

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	0.68±0.16	0.62±0.14
S5	0.79±0.21	0.69±0.18
Both	**0.70±0.17***	0.65±0.15

Declipped Using RBAR

- **Galaxy S5s perform better** than S3s when considering mania
  - Higher quality recordings
  - Subject population could also be more homogeneous
- **Significantly improved manic performance**
  - Significantly more clipping in manic calls than depressed calls from the S5
  - We hypothesize this is due to individuals speaking louder in a manic state

CHAI Lab
University of Michigan

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

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

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±0.24***	**0.77±0.15***

No Speech Segmentation

- **Segments were no longer found** using previous algorithm
  - All audio divided into 2 second segments with 1 second overlap
  - Results in much silence being captured
- Performs the **best of all tests**
  - Hypothesize this is actually caused by rhythm features **indirectly capturing information** about the assessment interview
  - Shows need for **accurate segmentation to avoid misleading results**

CHAI Lab
University of Michigan

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

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Results – Normalization By Subject

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	0.66±0.15	0.73±0.15
S5	0.71±0.35	0.78±0.10
Both	**0.67±0.19***	**0.75±0.14***

Normalized By Subject

- **Significant improvement for both mood tests**
- Previously shown to be able to correct for variations in feature distributions between speakers
  - Method also has ability to correct for phone models

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

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Results – Multi-Task Learning

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	0.67±0.20	0.67±0.21
S5	0.72±0.41	0.65±0.11
Both	**0.68±0.23***	0.66±0.18

Multi-Task Using Baseline Preprocessing

Model	Manic AUC	Depressed AUC
S3	0.71±0.19	0.66±0.14
S5	0.78±0.23	0.79±0.13
Both	**0.72±0.20***	0.71±0.15

Multi-Task Using Best Preprocessing

- **Significantly improves manic** test performance without any preprocessing modifications
- We hypothesize depressed tests are less affected due to being **more comparable before preprocessing**
- Best manic performance when using **all techniques**

CHAI Lab
University of Michigan

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

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN