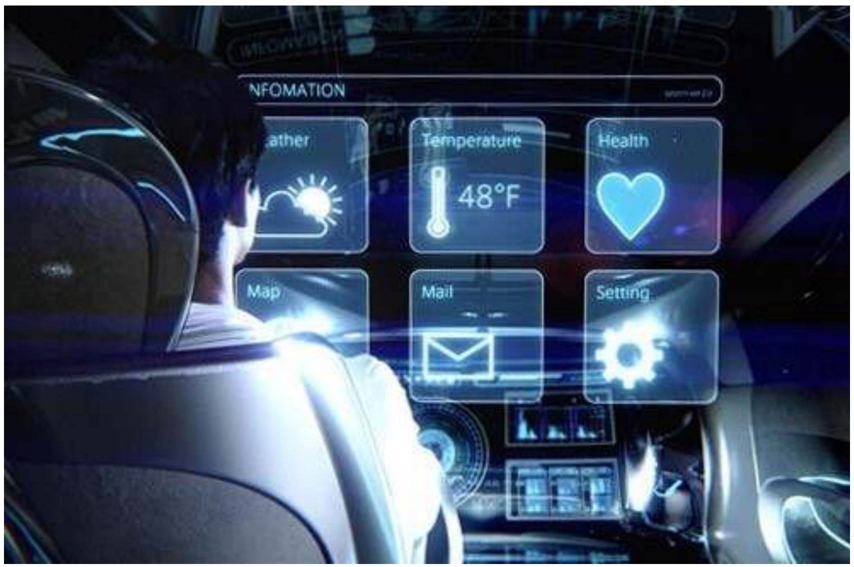
Human-Centered Computing: Using Speech to Understand Behavior

Emily Mower Provost

Computer Science and Engineering
University of Michigan

Automated agents provide support, entertainment, and interaction

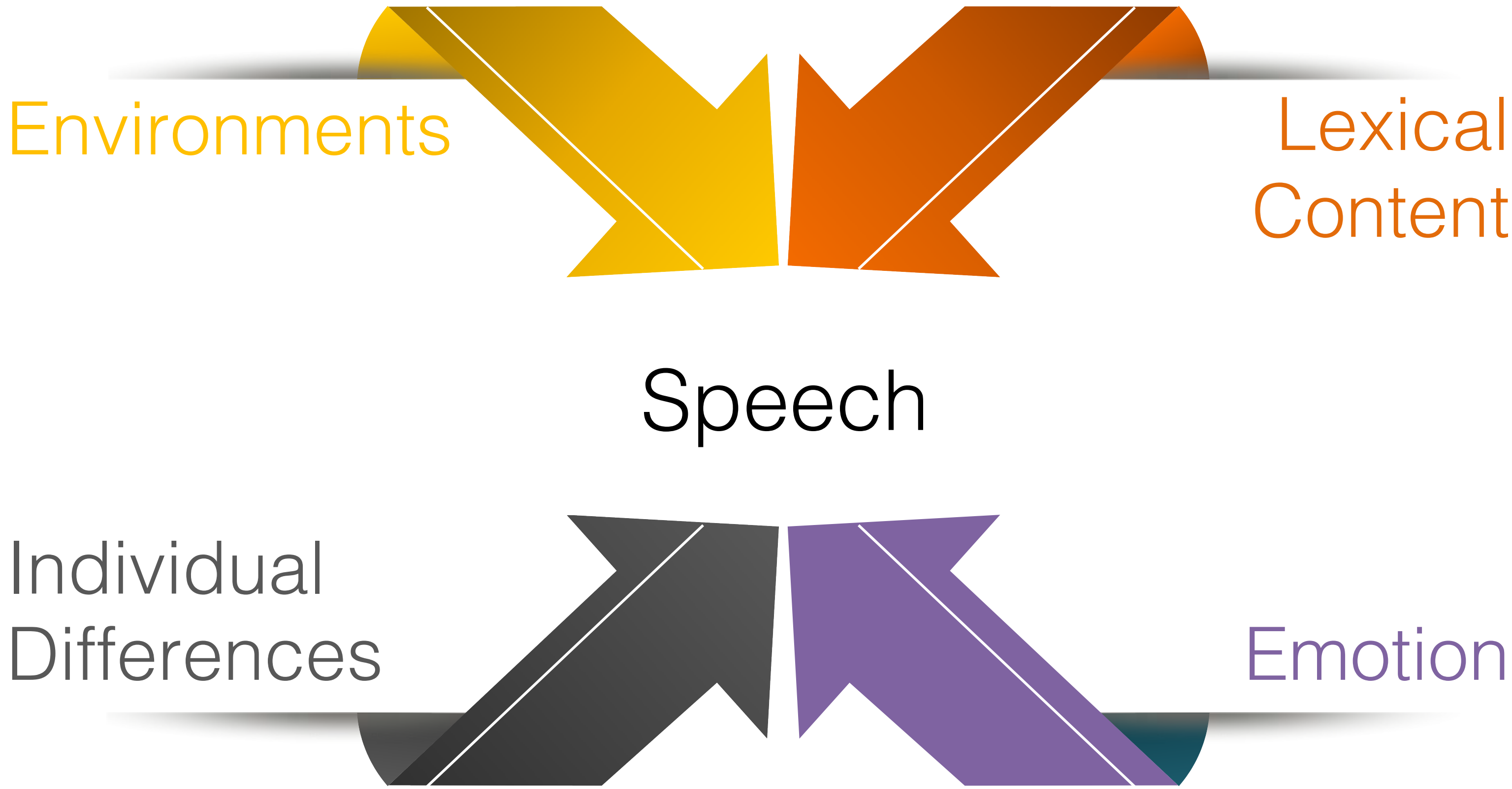
https://www.nimh.nih.gov/about/organization/gmh/grandchallenges/index.shtml



http://the-big-turn-on.co.uk/pics/future.jpg

Image source: http://newatlas.com/toyota-kirobo-mini-companion-robot-release/45720/

Embracing Complexity



CHAI Lab Research Directions

- Audio-visual emotion modeling:
  - Perception modeling
  - Expression modeling
  - Methods: deep learning, multitask learning, time series modeling, knowledge-driven

- Assistive technology:
  - Speech assessment for individuals with aphasia
  - Mood state tracking for individuals with bipolar disorder
  - [Early states] Estimating suicidality
  - [Early states] Speech assessment: Huntington's Disease

CHAI Lab
University of Michigan

COMPUTER SCIENCE
& ENGINEERING
UNIVERSITY OF MICHIGAN

CHAI Lab Research Directions

- Audio-visual emotion modeling:
  - Perception modeling
  - Expression modeling
  - Methods: deep learning, multitask learning, time series modeling, knowledge-driven

- Assistive technology:
  - Speech assessment for individuals with aphasia
  - Mood state tracking for individuals with bipolar disorder
  - [Early states] Estimating suicidality
  - [Early states] Speech assessment: Huntington's Disease

CHAI Lab Research Directions

- Audio-visual emotion modeling:
  - Perception modeling
  - Expression modeling
  - Methods: deep learning, multitask learning, time series modeling, knowledge-driven
- Assistive technology:
  - Speech assessment for individuals with aphasia
  - Mood state tracking for individuals with bipolar disorder
  - [Early states] Estimating suicidality
  - [Early states] Speech assessment: Huntington's Disease

Focus on Behaviors

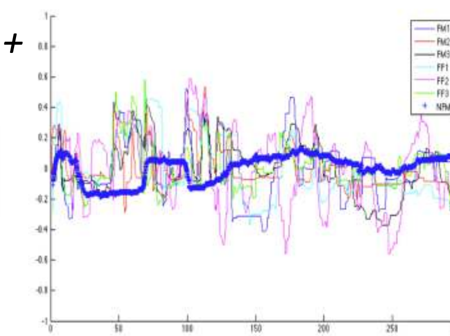
- Goal: detect human behavior from speech
  - Emotion: valence (positivity), activation (energy), categories
  - Mood: depression, suicidality
  - Diagnosis: Huntington Disease, aphasia

conversation

extract speech signal

speech signal

extract features + model



activation/valence patterns

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Why is this area so important?

ALGORITHMS → IMPACT

Motivation

- Bipolar Disorder (BP)
  - A leading cause of disability worldwide
  - Common, chronic, and severe psychiatric illness
  - Characterized by swings into mania and depression
  - Devastating personal, social, vocational consequences

Costly / Majority Unnecessary

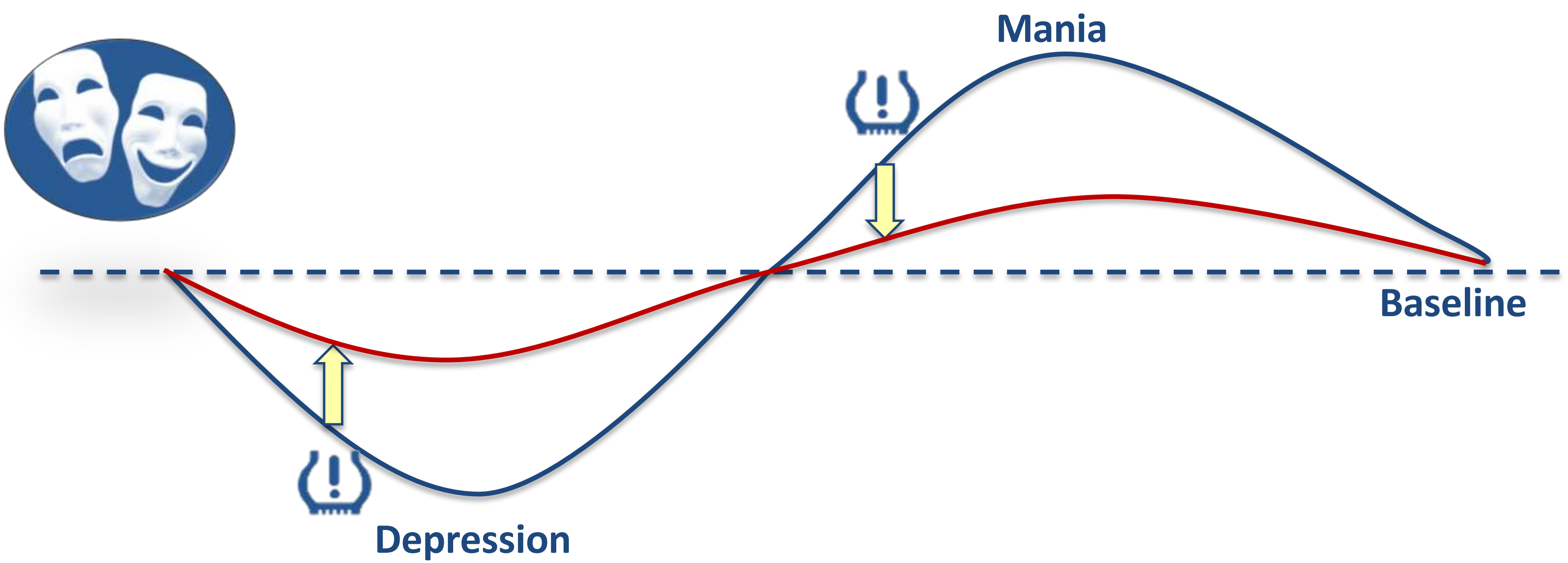
- Current Treatment
  - Pharmaceutically
  - Periodic follow-up visits for monitoring
  - Reactively post manic/depressive episodes

Devastating Consequences

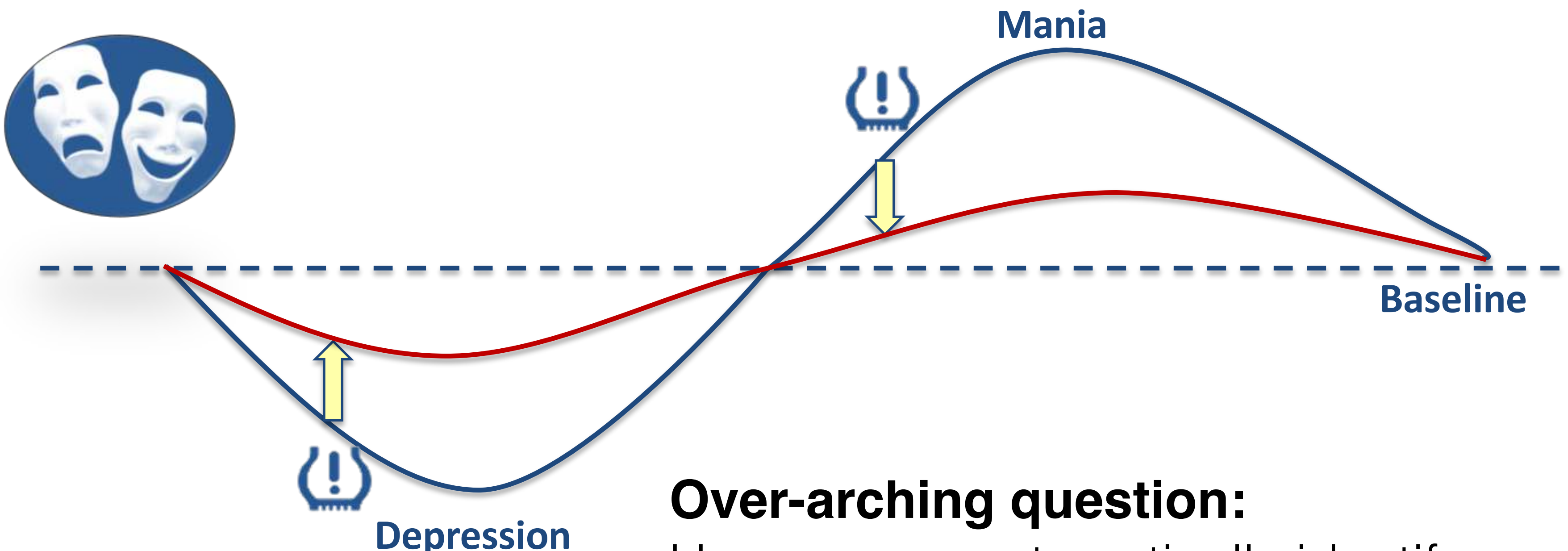
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Wellness Monitoring

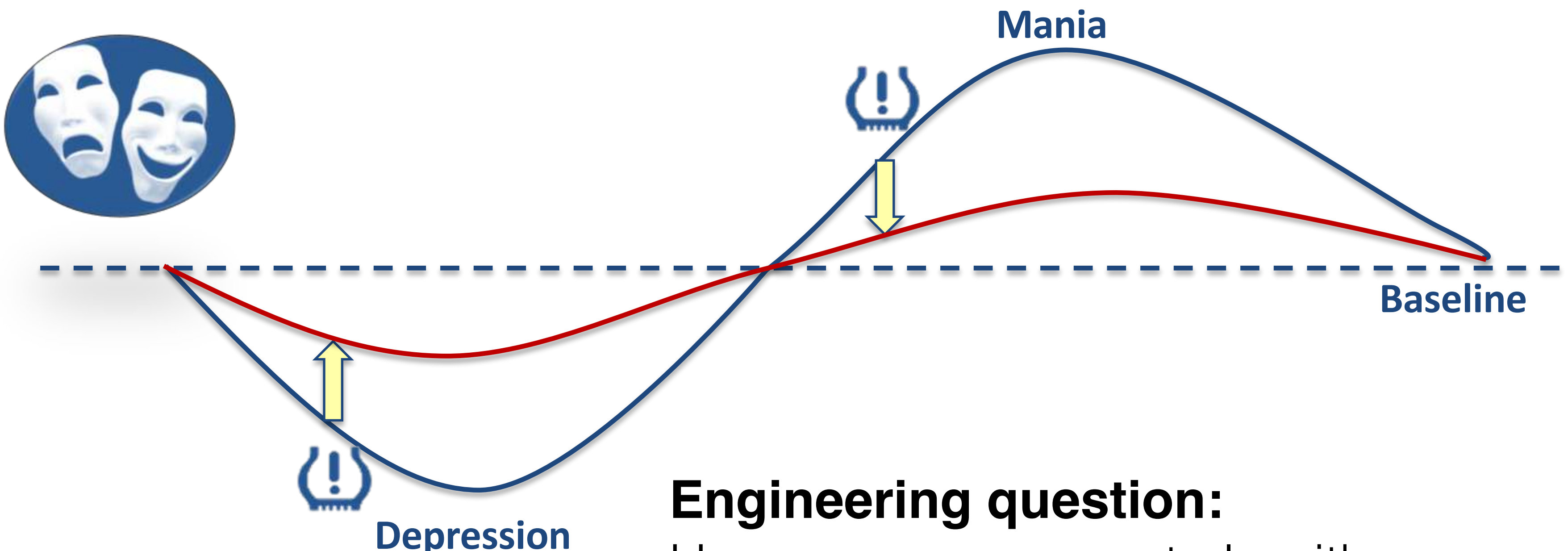


Wellness Monitoring



Over-arching question:
How can we automatically identify an individual's early warning signs?

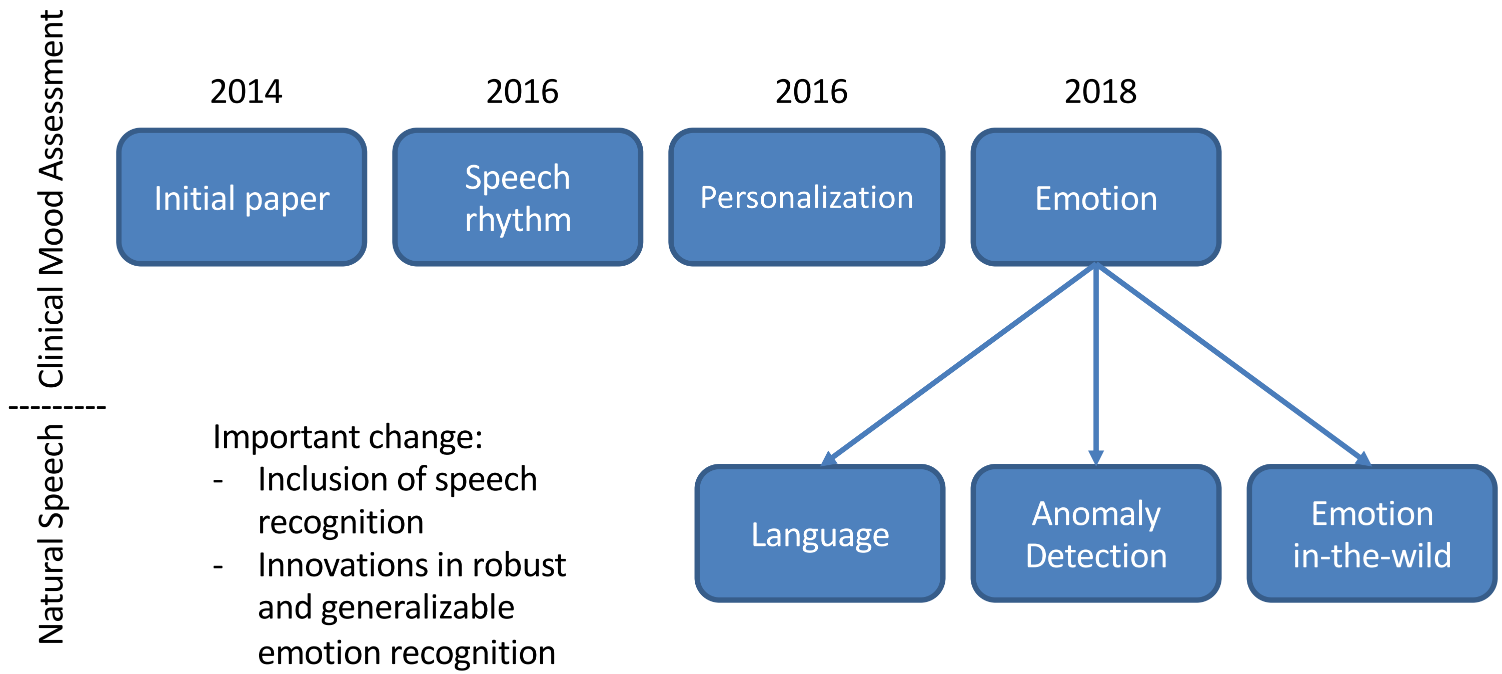
Wellness Monitoring



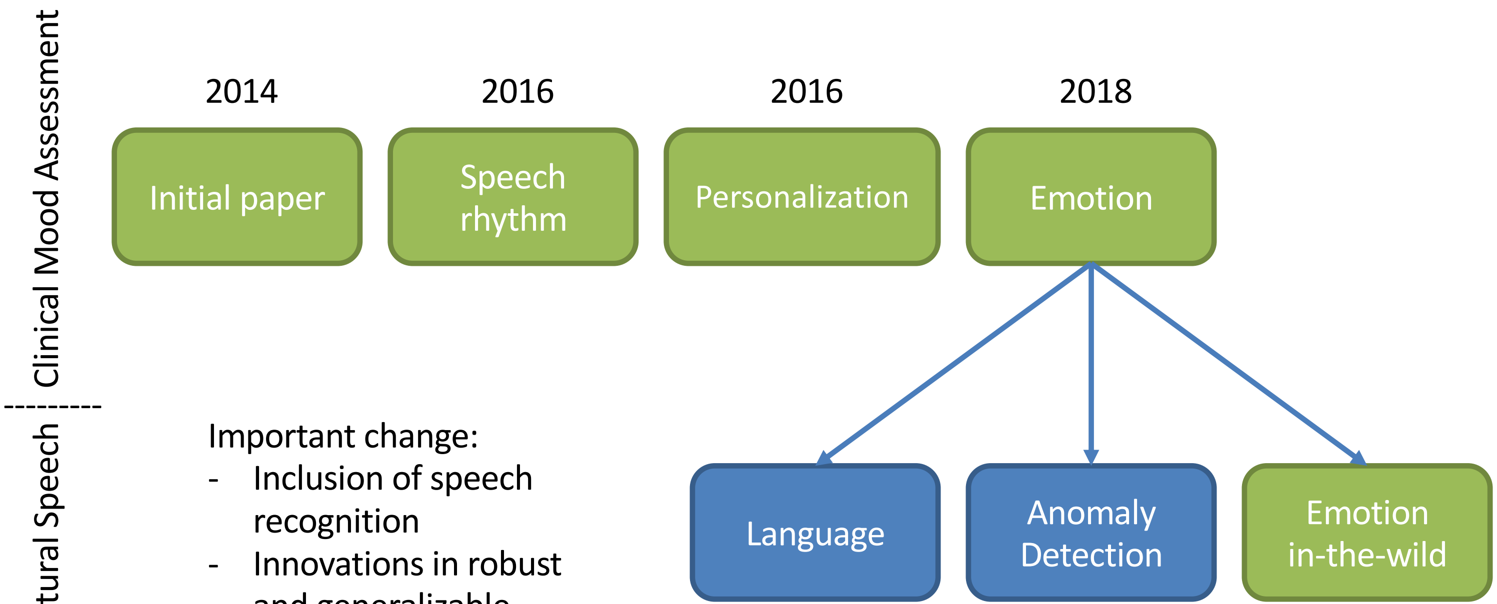
Engineering question:
How can we augment algorithm design with clinical knowledge?

This work:
What if we focus on emotion?

PRIORI Roadmap



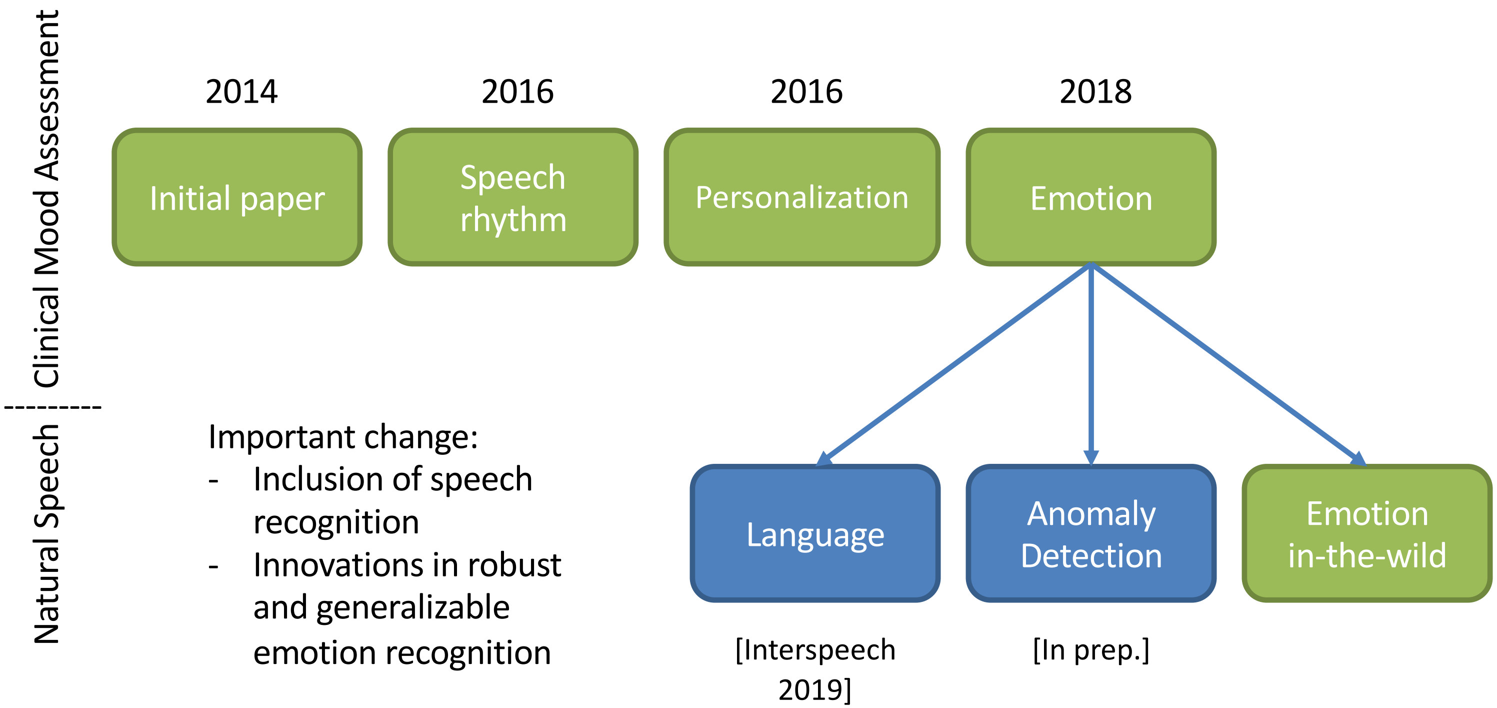
PRIORI Roadmap



Important change:

- Inclusion of speech recognition
- Innovations in robust and generalizable emotion recognition

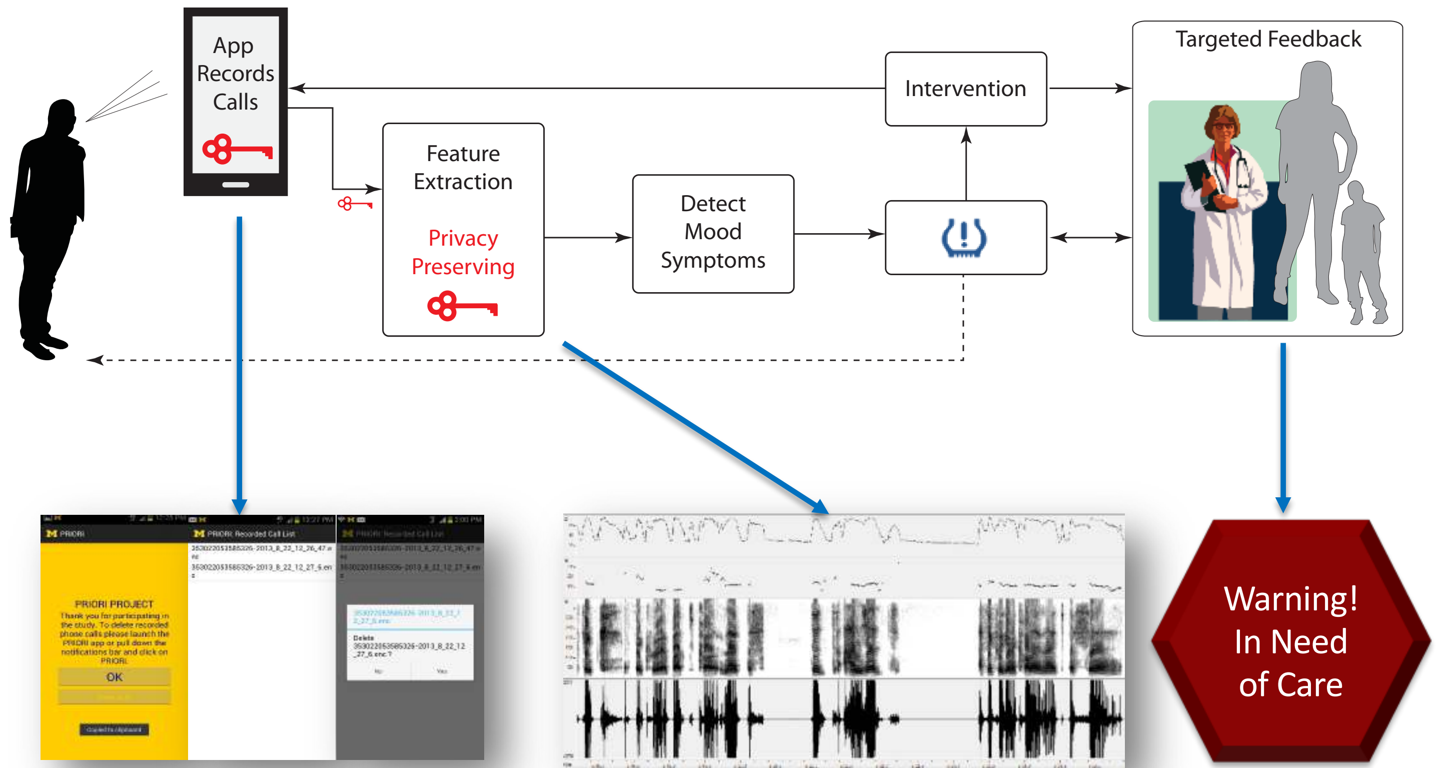
PRIORI Roadmap



Important change:

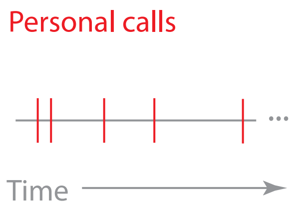
- Inclusion of speech recognition
- Innovations in robust and generalizable emotion recognition

PRedicting Individual Outcomes for Rapid Intervention



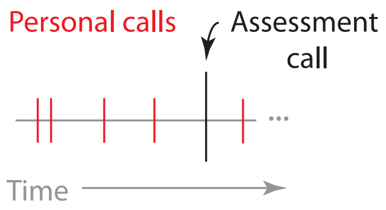
Types of PRIORI Calls

- Personal calls:
  - Calls made as someone goes about his/her day
  - Natural speech



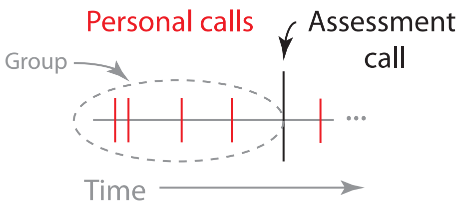
Types of PRIORI Calls

- Personal calls:
  - Calls made as someone goes about his/her day
  - Natural speech
- Assessment calls:
  - Clinical interactions over the phone
  - Young Mania Rating Scale (YMRS)
  - Hamilton Depression Scale (HamD)



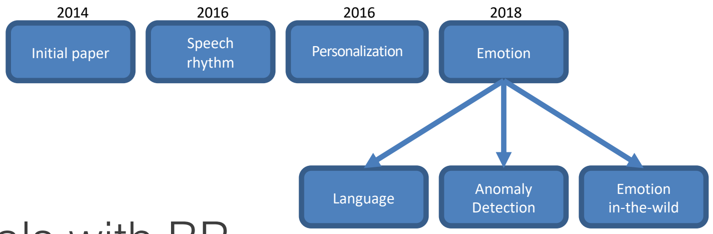
Types of PRIORI Calls

- Personal calls:
  - Calls made as someone goes about his/her day
  - Natural speech
- Assessment calls:
  - Clinical interactions
  - Young Mania Rating Scale (YMRS)
  - Hamilton Depression Scale (HamD)



Personal calls **grouped** by assessment call

The PRIORI dataset



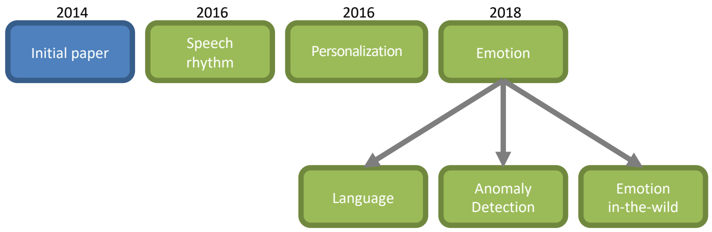
- PRIORI:
  - Longitudinal study of bipolar disorder
  - Collect and analyze mood data for individuals with BP
  - Develop a mood recognition systems

- Participants
  - Patients: BP I and II (51)
  - Healthy controls (9)
  - Dataset size: over 50K calls, over 4K hours of speech

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Initial Paper



Goal
- Collect and present a new dataset!
- Determine the efficacy of emotion techniques for recognizing mood

Insight
- Emotion and mood both modulate the speech signal

Approach
- Extract common emotion features
- Classify using common emotion recognition techniques

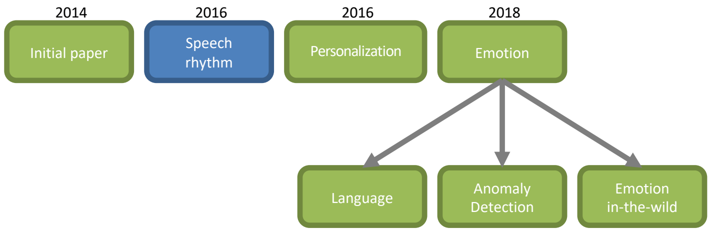
Findings
- There do seem to be differences!

CHAI Lab
University of Michigan

Zahi N Karam, Emily Mower Provost, Satinder Singh, Jennifer Montgomery, Christopher Archer, Gloria Harrington, Melvin Mcinnis. ``Ecologically Valid Long-term Mood Monitoring of Individuals with Bipolar Disorder Using Speech.'' International Conference on Acoustics, Speech and Signal Processing (ICASSP). Florence, Italy. May 2014.

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Speech Rhythm



Goal
- Determine whether a clinician would designate a person in a mood episode using the rhythm of speech in a clinical interaction

Insight
- When manic, speech rate increases, when depressed, it decreases

Approach
- Create a robust pre-processing pipeline
- Classify mood episode

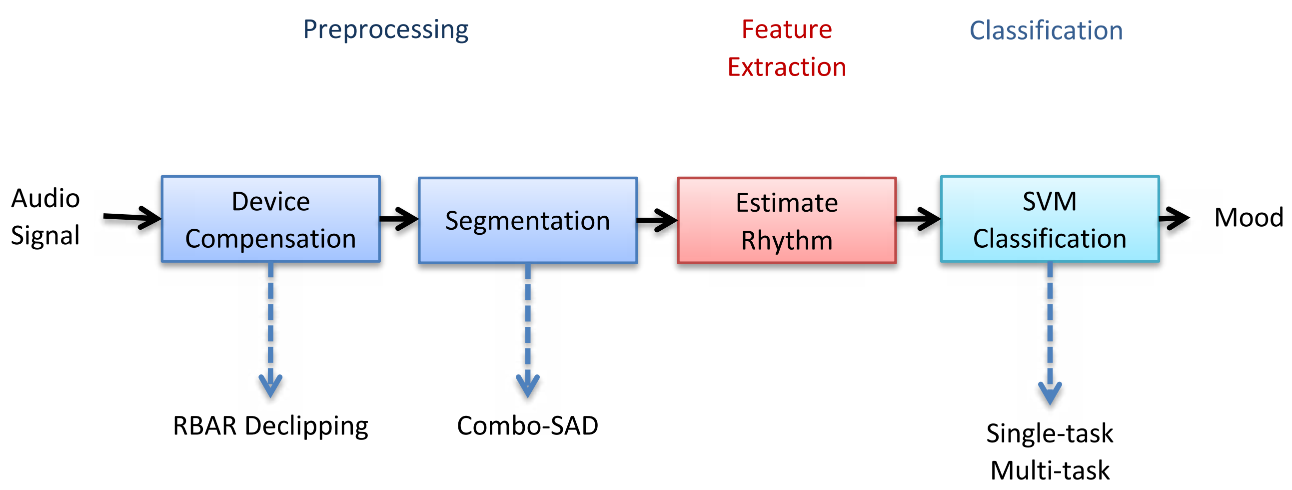
Findings
- Rhythm can be used to estimate mood
- It is critical to control for extraneous factors!

CHAI Lab University of Michigan

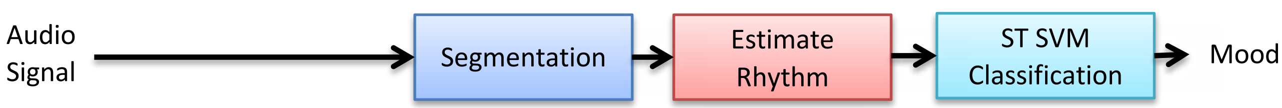
John Gideon, Emily Mower Provost, Melvin McInnis. “Mood State Prediction From Speech Of Varying Acoustic Quality For Individuals With Bipolar Disorder.” International Conference on Acoustics, Speech and Signal Processing (ICASSP). Shanghai, China, March 2016.

COMPUTER SCIENCE & ENGINEERING UNIVERSITY OF MICHIGAN

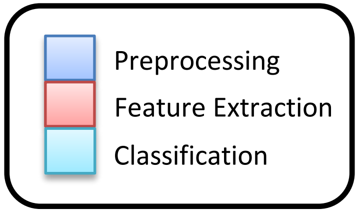
Methods


Preprocessing
Feature
Extraction
Classification
Audio
Signal
Device
Compensation
Segmentation
Estimate
Rhythm
SVM
Classification
Mood
RBAR Declipping
Combo-SAD
Single-task
Multi-task

Methods and Results



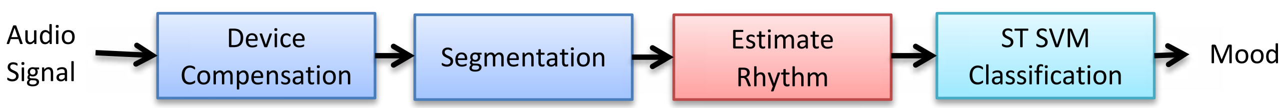
Measure: AUC	Baseline
Mania	0.57 ± 0.25
Depression	0.64 ± 0.14



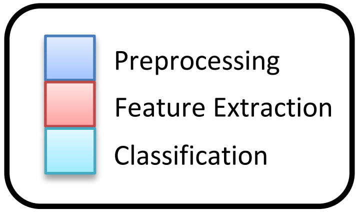
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Methods and Results

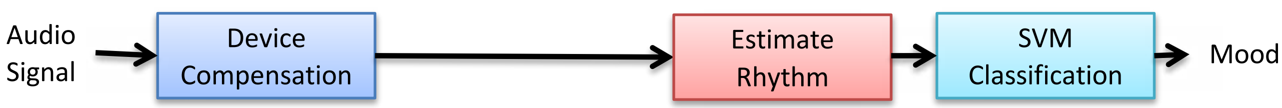


Measure: AUC	Baseline	RBAR Declipping
Mania	0.57 ± 0.25	0.70 ± 0.17*
Depression	0.64 ± 0.14	0.65 ± 0.15

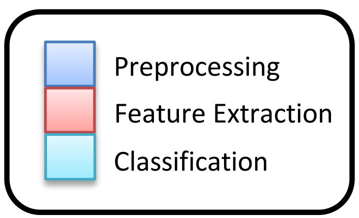


* Paired t-test over subjects, $p < 0.05$

Methods and Results

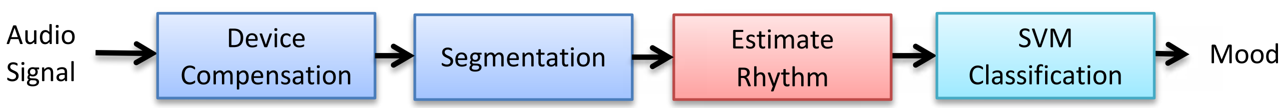


Measure: AUC	Baseline	RBAR Declipping	Ignoring Segmentation
Mania	0.57 ± 0.25	0.70 ± 0.17*	0.74 ± 0.24*
Depression	0.64 ± 0.14	0.65 ± 0.15	0.77 ± 0.15*

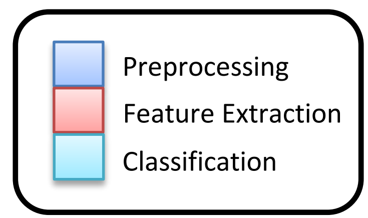


* Paired t-test over subjects, $p < 0.05$

Methods and Results

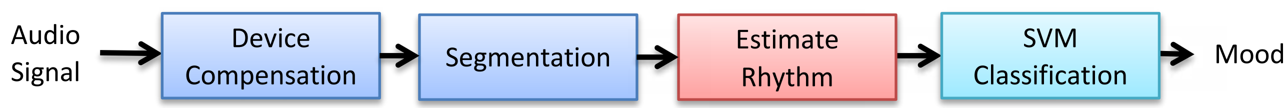


Measure: AUC	Baseline	RBAR Declipping	Multitask Learning
Mania	0.57 ± 0.25	0.70 ± 0.17*	0.72 ± 0.20*
Depression	0.64 ± 0.14	0.65 ± 0.15	0.71 ± 0.15



* Paired t-test over subjects, $p < 0.05$

Methods and Results

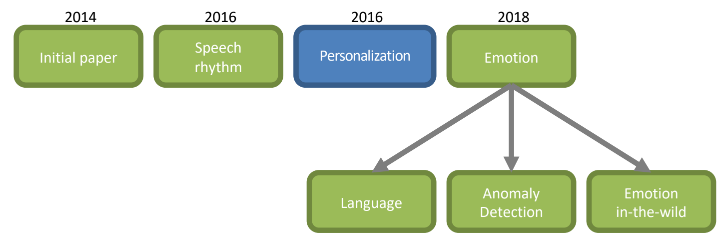


Measure: AUC	Baseline	RBAR Declipping	Subject Normalization
Mania	0.57 ± 0.25	0.70 ± 0.17*	0.67 ± 0.19*
Depression	0.64 ± 0.14	0.65 ± 0.15	0.75 ± 0.14*

Preprocessing
Feature Extraction
Classification

* Paired t-test over subjects, $p < 0.05$

Personalization



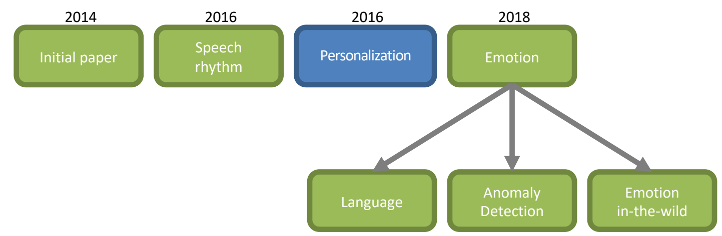
Goal	• Improve the prediction of depression
Insight	• Individuals are unique and so is their expression of mood
Approach	• Speaker verification techniques (i-vectors)
Findings	• We can improve depression prediction over speech rhythm features alone

Soheil Khorram, John Gideon, Melvin McInnis, and Emily Mower Provost. "Recognition of Depression in Bipolar Disorder: Leveraging Cohort and Person-Specific Knowledge." Interspeech. San Francisco, CA, September 2016.

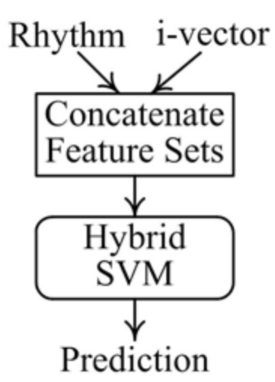
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

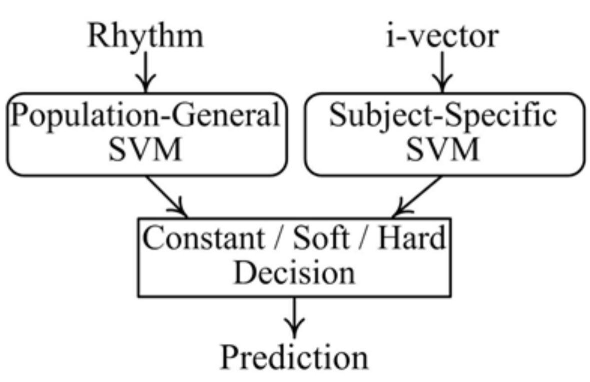
Personalization



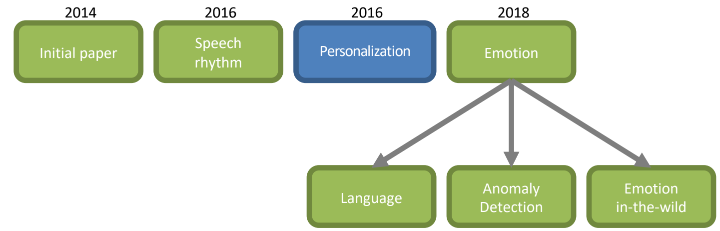
Feature Fusion



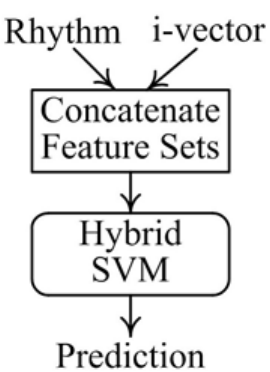
Decision Fusion



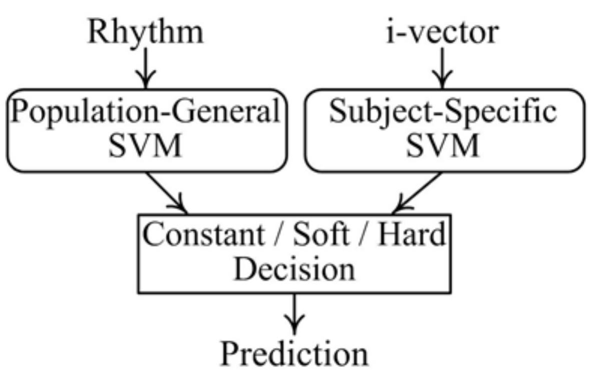
Personalization



Feature Fusion



Decision Fusion



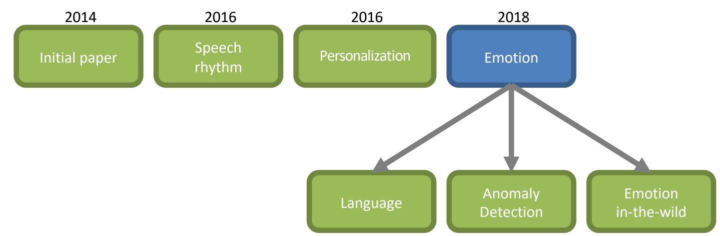
Results

System Characteristics	AUC
Population-general	0.69 ± 0.15
Subject-specific	0.70 ± 0.18
Feature Fusion	**0.76 ± 0.13***
Constant Decision Fusion	0.74 ± 0.16
Soft Decision Fusion	**0.78 ± 0.12***
Hard Decision Fusion	0.76 ± 0.13

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Emotion



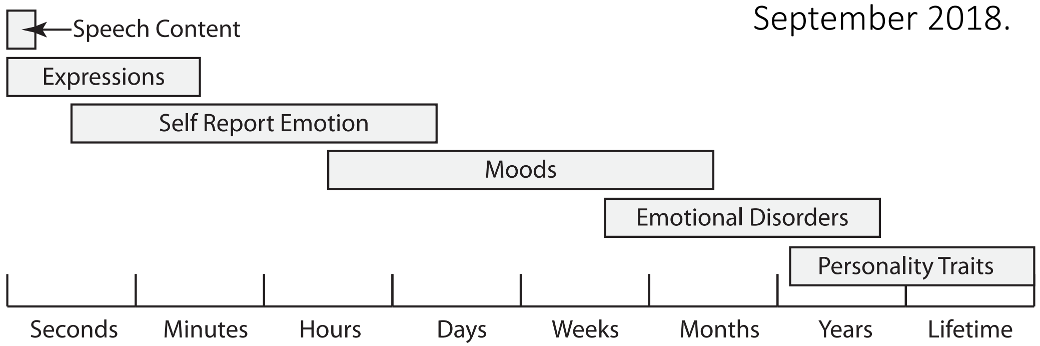
Goal	• Move to personal calls!
Insight	• Mood is slowly varying, can we improve prediction by focusing on factors more directly expressed in speech?
Approach	• Annotate the data for emotion! • Transcribe the data!
Findings	• We can accurately predict emotion from natural speech • In clinical interactions, emotion patterns change with symptom severity

Soheil Khorram, Mimansa Jaiswal, John Gideon, Melvin McInnis, Emily Mower Provost. "The PRIORI Emotion Dataset: Linking Mood to Emotion Detected In-the-Wild." Interspeech. Hyderabad, India. September 2018.

Identifying an intermediary step

- Mood prediction is challenging:
  - Not directly observable
  - Long time scale

Reference:
Soheil Khorram, Mimansa Jaiswal, John Gideon, Melvin McInnis, Emily Mower Provost. “The PRIORI Emotion Dataset: Linking Mood to Emotion Detected In-the-Wild.” Interspeech. Hyderabad, India. September 2018.



- Emotion can simplify mood prediction:
  - Primary BP symptom: emotion dysregulation, utility in classification*
  - Time course: emotion variation between speech and mood

CHAI Lab
University of Michigan

* Stasak, B., Epps, J., Cummins, N., & Goecke, R. (2016). An Investigation of Emotional Speech in Depression Classification. In INTERSPEECH (pp. 485-489).

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

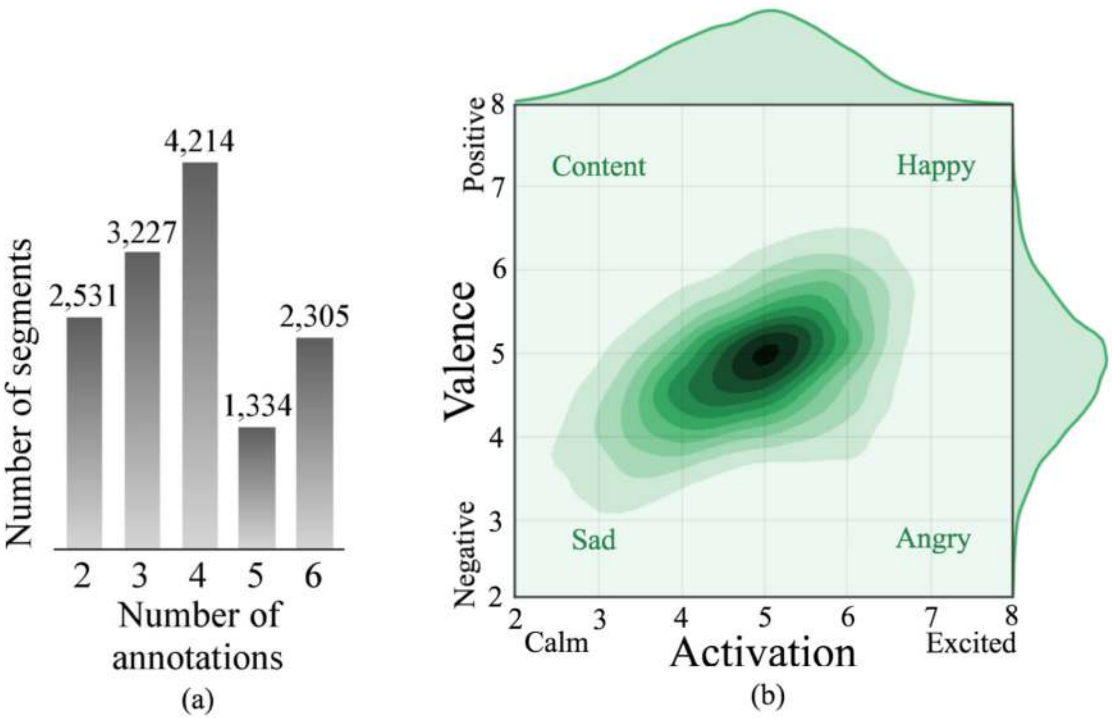
Emotion Annotation Pipeline

Subject Selection → Segmentation → Segment Selection → Segment Inspection → Segment Annotation

- Valence and activation annotation:
  - 9-point Likert scale
  - 11 annotators (7 female, 4 male), between 21 and 34, native speakers of English

- Annotators were asked to consider two important points:
  - Only the acoustic characteristics, not the content
  - Subject-specificity of emotion expression

Emotion Distributions



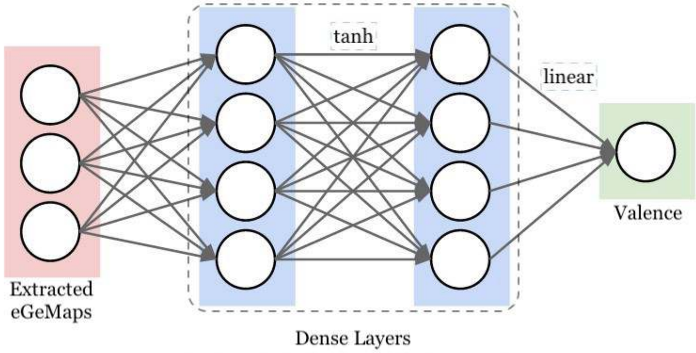
*Note: categorical labels for demonstration purposes only.

Emotion Recognition Experimental Setup

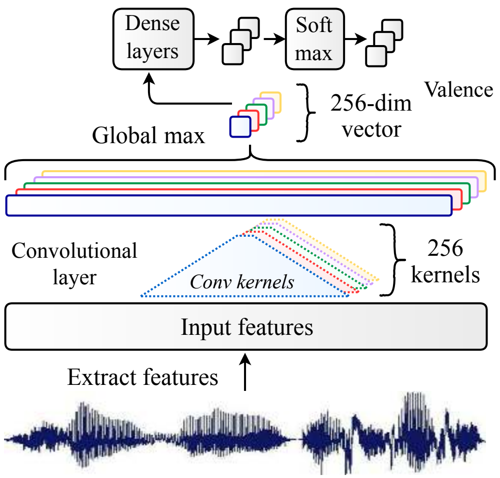
- Normalize ground truth labels:
  - Subtracting the rating midpoint of 5
  - Scaling to the range of [–1, 1]
- Subject-independent cross-validation
  - Experiments repeated for five total runs (six randomly selected folds)
  - Each run: randomly assign two subjects to each fold.
  - Round-robin cross-validation
  - Generates one test measure per fold, resulting in six measures.
  - Output: matrix of 6-by-5 test measures
- Parameter selection: max CCC over validation set

Features and Models

- Baseline system
  - 88-dimensional eGeMAPS features
  - Features globally normalized
  - Feed-forward neural network,
    tanh activation function, linear output



- Alternative system
  - 40-dimensional MFB features
  - Features globally normalized
  - Conv-pool network (convolutional layers, global max pooling, dense layers)
  - ReLU and linear activation functions for intermediate and output



Emotion Results

- Conv-Pool > FFNN (PCC, CCC)

Dimension	Metric	eGeMAPS FFN	MFBs Conv-Pool
Activation	PCC	0.642 ± 0.076	**0.712 ± 0.077**
	CCC	0.593 ± 0.071	**0.660 ± 0.090**
	RMSE	0.207 ± 0.012	0.201 ± 0.028
Valence	PCC	0.271 ± 0.053	**0.405 ± 0.062**
	CCC	0.191 ± 0.031	**0.326 ± 0.052**
	RMSE	0.199 ± 0.015	0.194 ± 0.016

Bold: $p<0.01$, paired t-test

Emotion Results

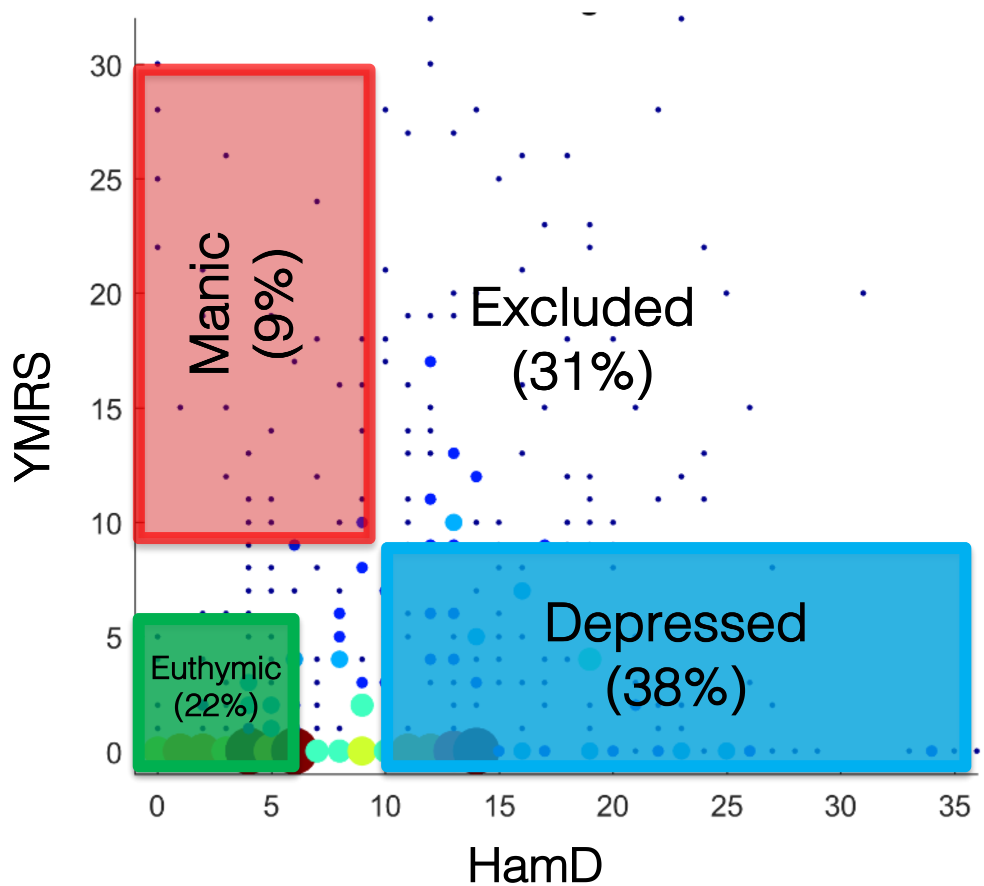
- Conv-Pool > FFNN (PCC, CCC)
- Activation more accurately recognized

Dimension	Metric	eGeMAPS FFN	MFBs Conv-Pool
Activation	PCC	0.642 ± 0.076	**0.712 ± 0.077**
	CCC	0.593 ± 0.071	**0.660 ± 0.090**
	RMSE	0.207 ± 0.012	0.201 ± 0.028
Valence	PCC	0.271 ± 0.053	**0.405 ± 0.062**
	CCC	0.191 ± 0.031	**0.326 ± 0.052**
	RMSE	0.199 ± 0.015	0.194 ± 0.016

Bold: $p<0.01$, paired t-test

Mood Dataset

- **Goal**: Analyze link between mood and predicted emotion



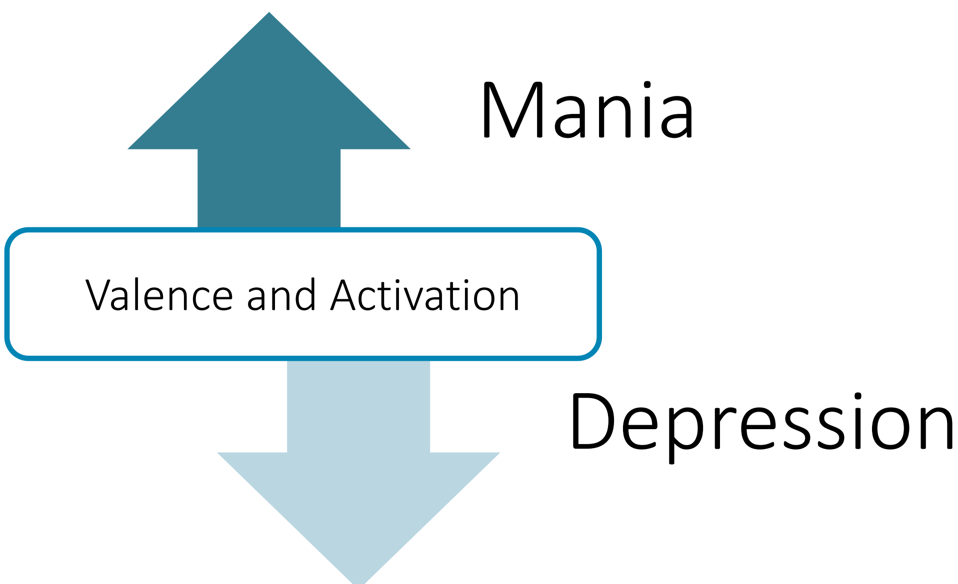
Experimental Setup

- **Goal**: Analyze link between mood and predicted emotion
- Considerations:
  - Importance of considering how a subject varies about his/her own baseline (euthymic periods)
  - Normalize depressed, manic segments by subject (euthymic segments)
- Approach:
  - Apply conv-pool models to predict emotion
  - Use ensemble over the cross-validation models
  - Analyze over all 10,563 assessment call segments (10,563)

What is the link between mood and emotion?

- Ways to measure:
  - Observe **clinical interactions**
  - Relate emotion to mood symptom severity (classes or continuous)

- **Finding**: valence/activation significantly **higher** in manic vs. depressed episodes

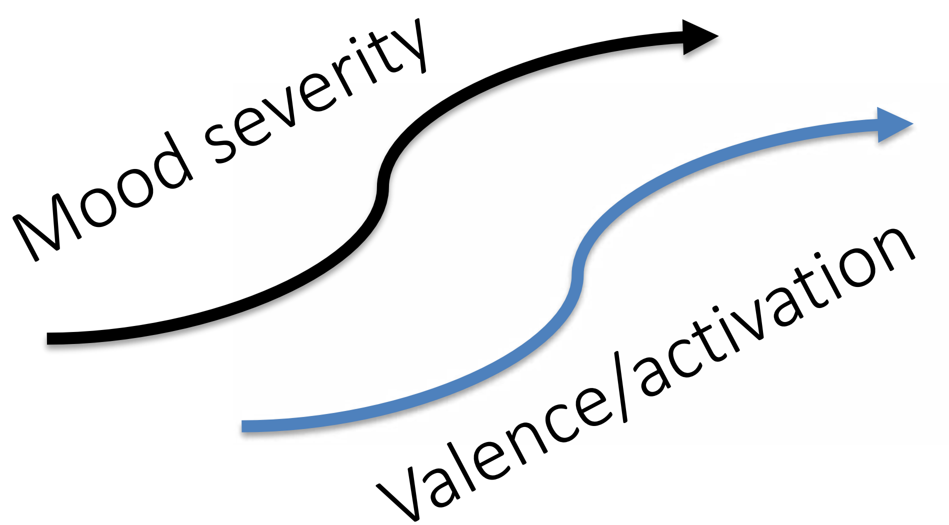


Valence: positive vs. negative
Activation: calm vs. excited

What is the link between mood and emotion?

- Ways to measure:
  - Observe **clinical interactions**
  - Relate emotion to mood symptom severity (classes or continuous)

- **Finding**: valence/activation are significantly **correlated** with mood severity

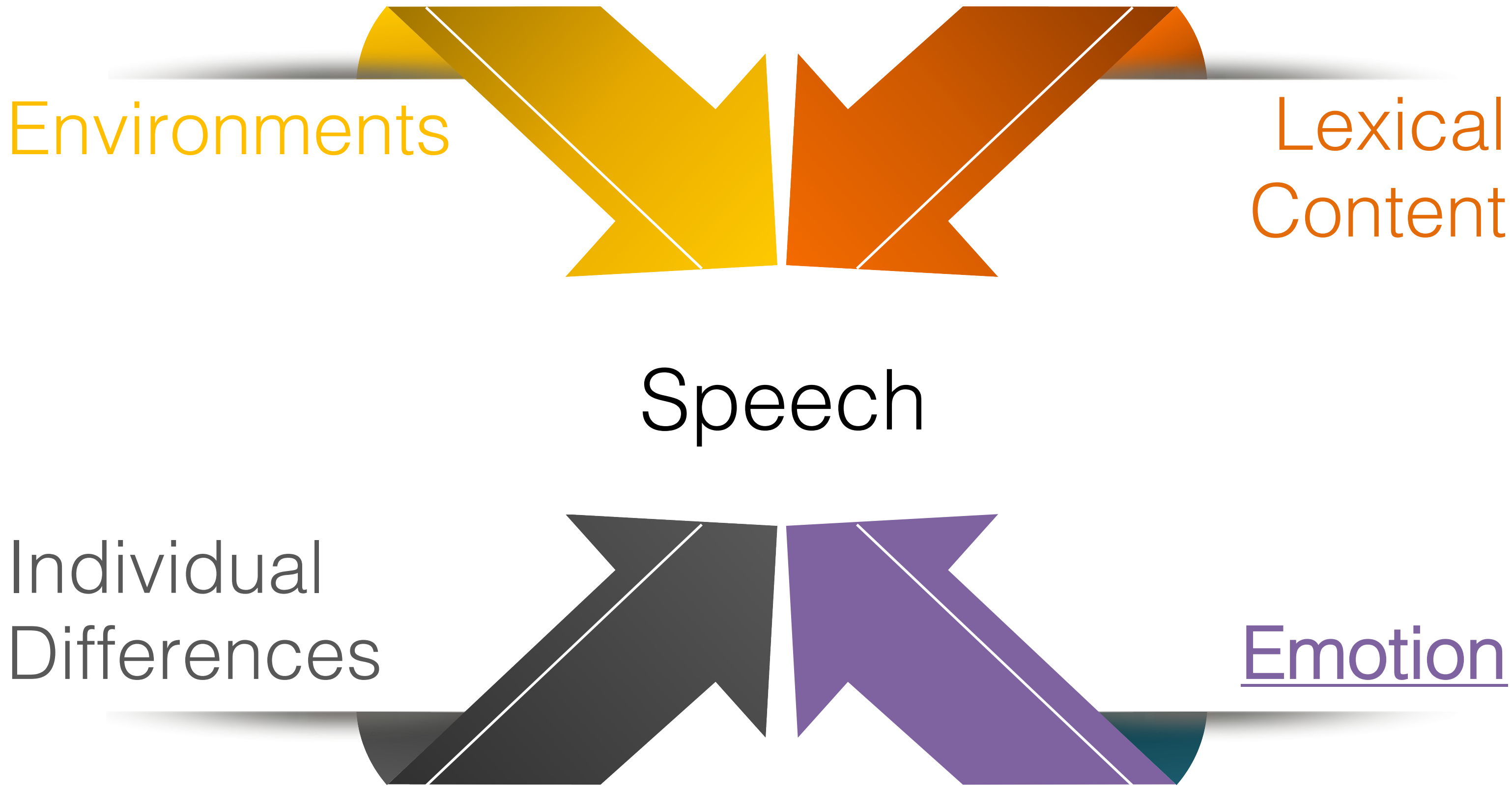


Valence: positive vs. negative
Activation: calm vs. excited

Comparing Emotion Distributions

- Comparing distributions of valence/activation across subjects
- Comparisons:
  - Over all subjects: one-way ANOVA with $p < 0.01$
  - Pairwise comparisons: Tukey-Kramer posthoc test (66 pairs)
- Findings:
  - Activation: overall difference, significantly different in 51 cases
  - Valence: overall difference, significantly different in 48 cases

Embracing Complexity



Research Question

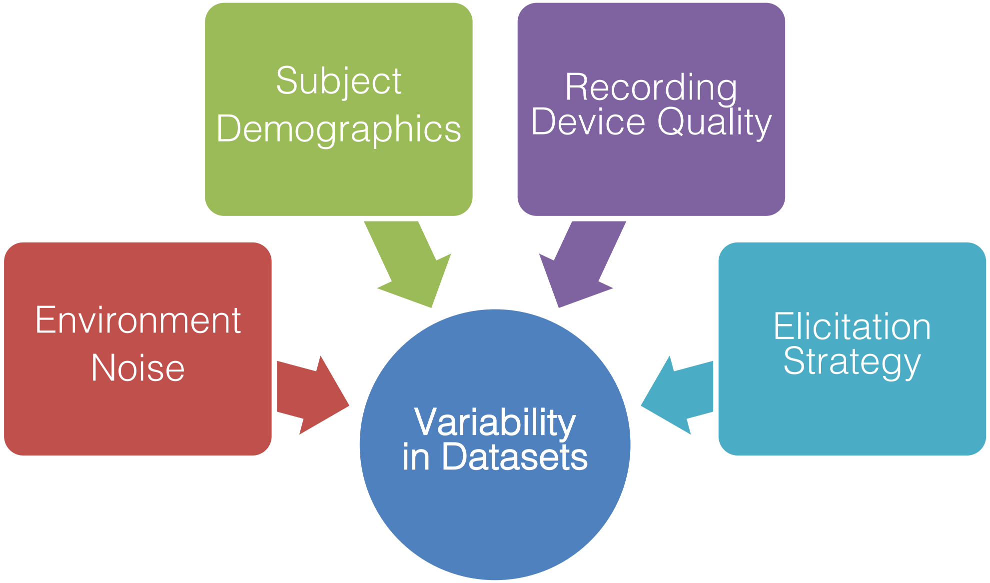
Emotion is a big data problem!

But, what is the best method for transferring **paralinguistic information** and **datasets with different conditions** to emotion?

Reference:
John Gideon, Melvin McInnis, Emily Mower Provost. "Barking up the Right Tree: Improving Cross-Corpus Speech Emotion Recognition with Adversarial Discriminative Domain Generalization (ADDoG)," IEEE Transactions on Affective Computing, vol: To appear, 2019.

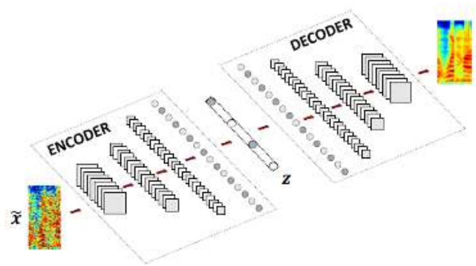
Domain Generalization

- Goal: creates a **middle-ground** representation for **unseen data**
- Removes factors particular to individual datasets

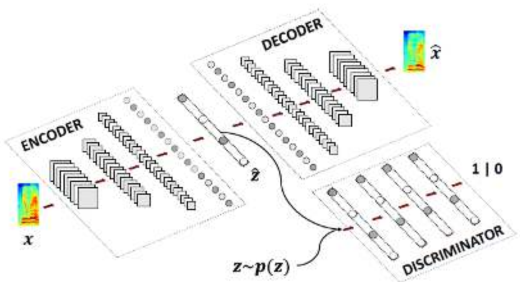


Domain Generalization – Autoencoders

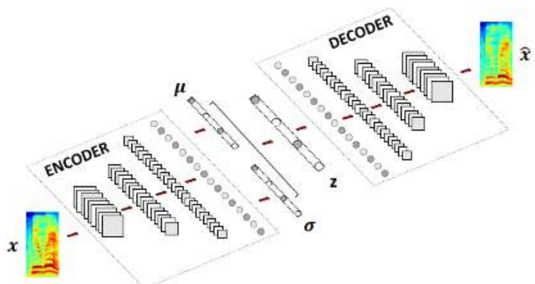
Denoising Autoencoder (DAE)



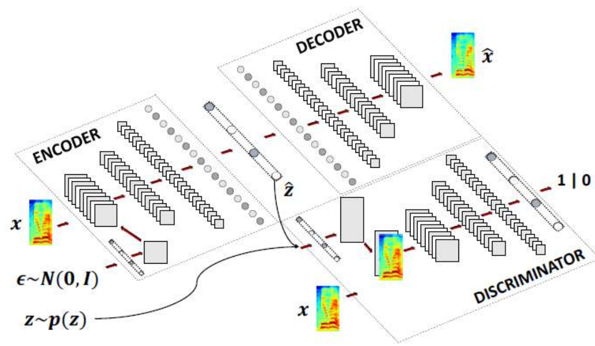
Adversarial Autoencoder (AAE)



Variational Autoencoder (VAE)



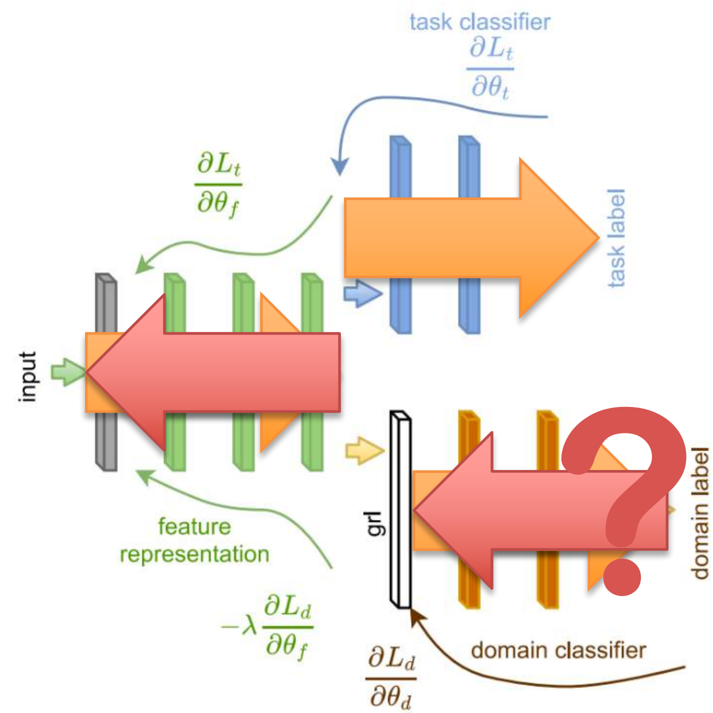
Adversarial Variational Bayes (AVB)



Eskimez et al. 2018

Domain Generalization – DANNs

- Domain Adversarial Neural Networks
- **Encode** a middle representation
- **Discriminative**: Classify emotion and domain from middle layer
- **Adversarial**: Backpropagate the reverse gradient of domain
- "**Unlearns**" domain
- No clear target – challenges with converging



Ajakan et al. 2014; Abdelwahab et al. 2018

What if we could still be discriminative?

What if we could still be discriminative?

What if we had a clear target?

CHAI Lab
University of Michigan

COMPUTER SCIENCE
& ENGINEERING
UNIVERSITY OF MICHIGAN

Datasets

	IEMOCAP	MSP-IMPROV
Subjects (Male/Female)	10 (5/5)	12 (6/6)
Environment	Laboratory	Laboratory
Language	English	English
Sample Rate	16 kHz	44.1 kHz
Total Utterances	10039	8438

Labels

	IEMOCAP	MSP-IMPROV
Total Utterances	10039	8438
Likert Scale	1-5	1-5
Class Boundaries	1-2, 3, 4-5	1-2, 3, 4-5
Mean (Std.) Activation	3.08 (0.90)	2.57 (1.10)
Utt. Without Ties	4814	7290
Mean (Std.) Valence	2.79 (0.99)	3.02 (1.06)
Utt. Without Ties	6816	7852

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Labels

	IEMOCAP	MSP-IMPROV
Total Utterances	10039	8438
Likert Scale	1-5	1-5
Class Boundaries	1-2, 3, 4-5	1-2, 3, 4-5
Mean (Std.) Activation	3.08 (0.90)	2.57 (1.10)
Utt. Without Ties	4814	7290
Mean (Std.) Valence	2.79 (0.99)	3.02 (1.06)
Utt. Without Ties	6816	7852

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

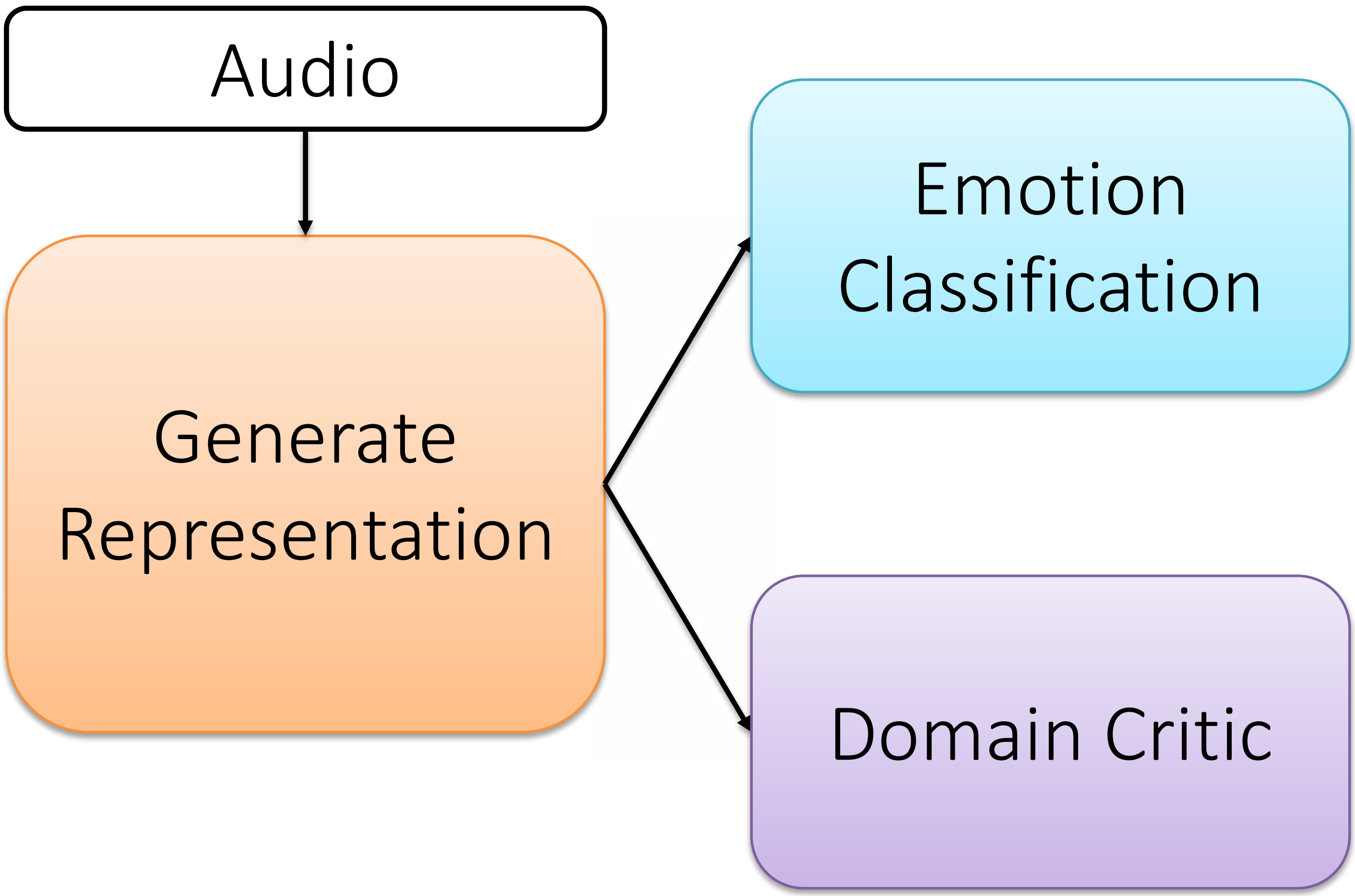
Labels

	IEMOCAP	MSP-IMPROV
Total Utterances	10039	8438
Likert Scale	1-5	1-5
Class Boundaries	1-2, 3, 4-5	1-2, 3, 4-5
Mean (Std.) Activation	3.08 (0.90)	2.57 (1.10)
Utt. Without Ties	4814	7290
Mean (Std.) Valence	2.79 (0.99)	3.02 (1.06)
Utt. Without Ties	6816	7852

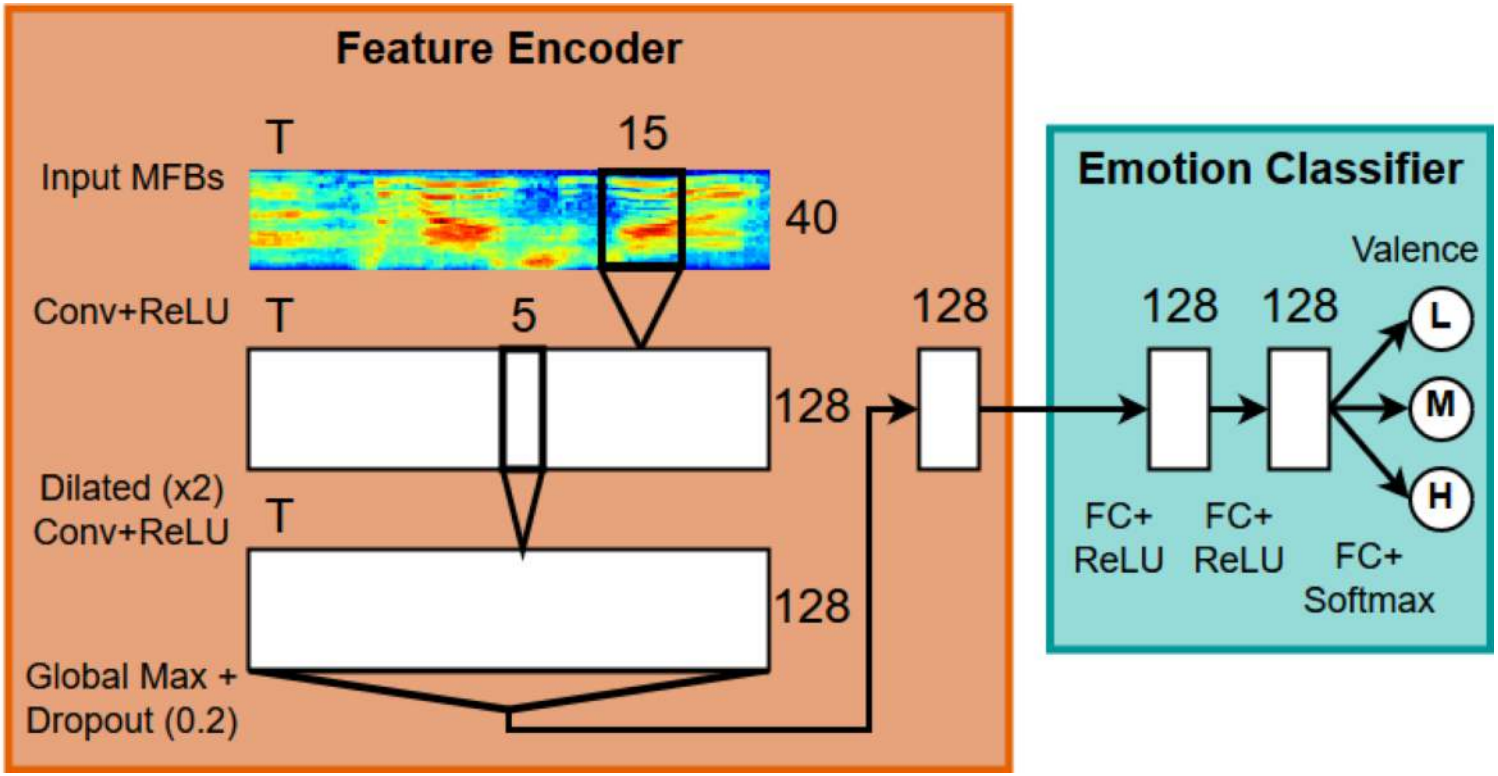
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Method Overview

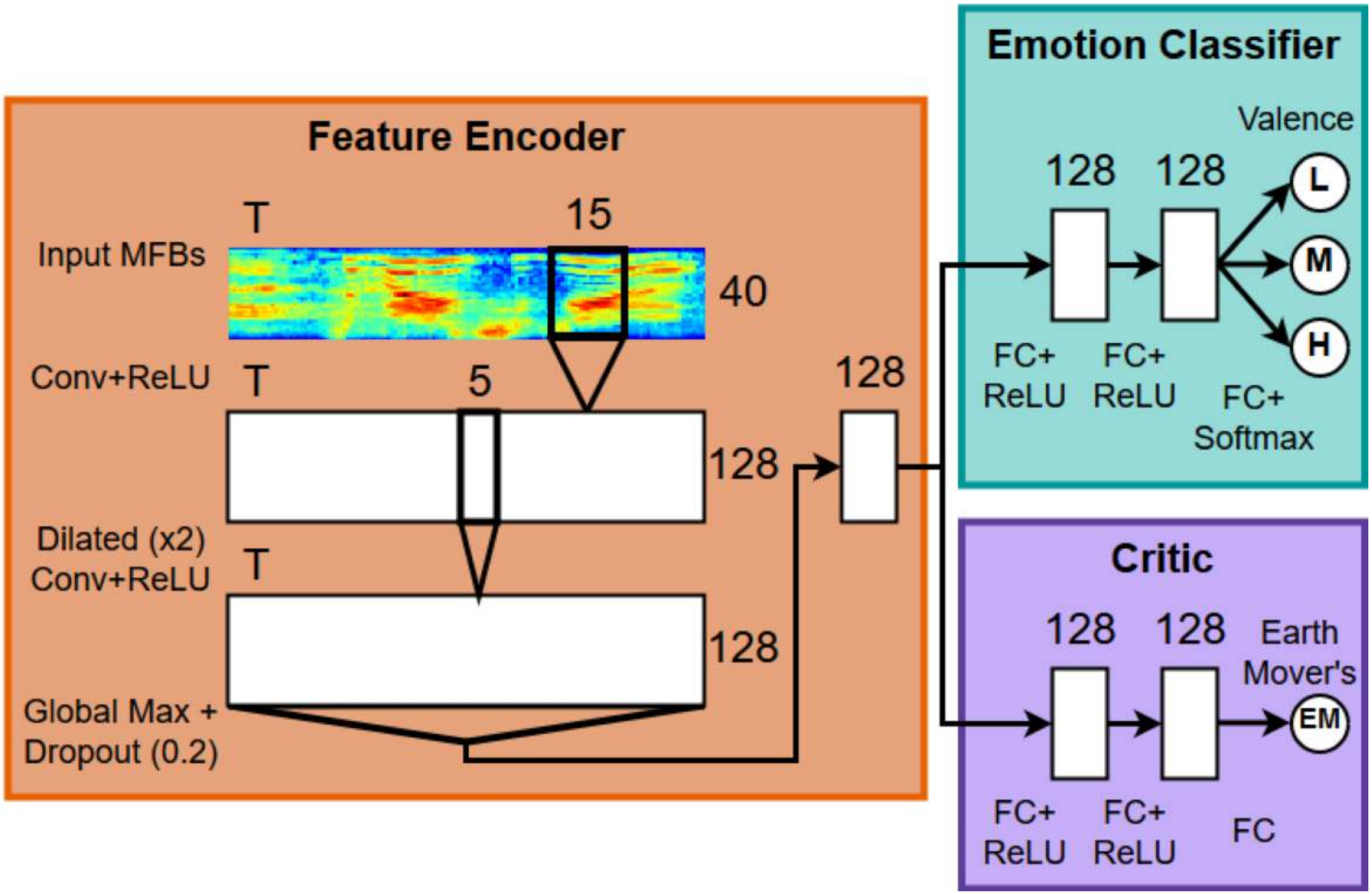


Baseline: CNN

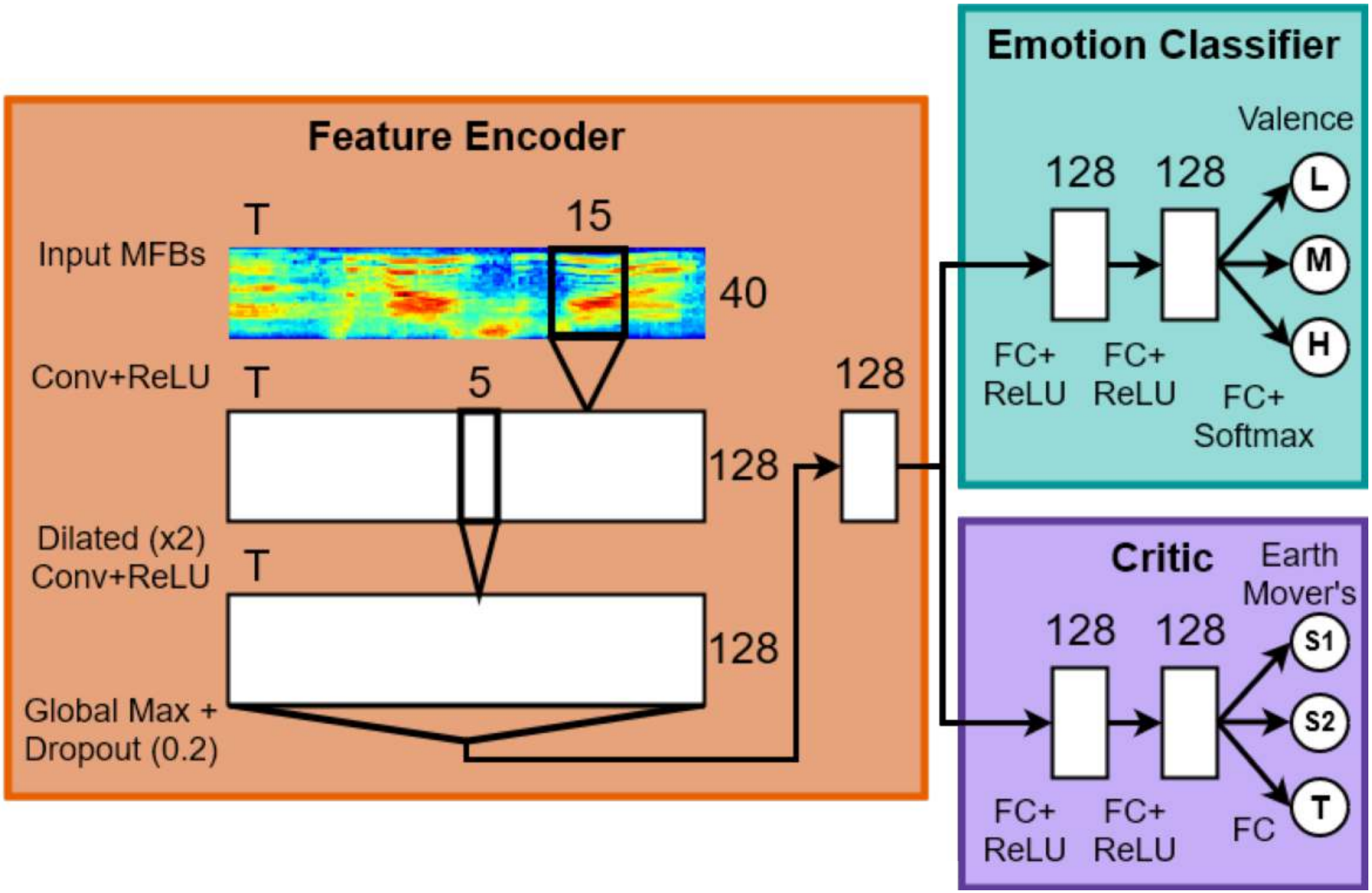


CNN: Convolutional Neural Network trained on all labeled data;
SP: Specialist CNN trained on just target labeled data (if available)

ADDoG: Adversarial Discriminative Domain Gen.



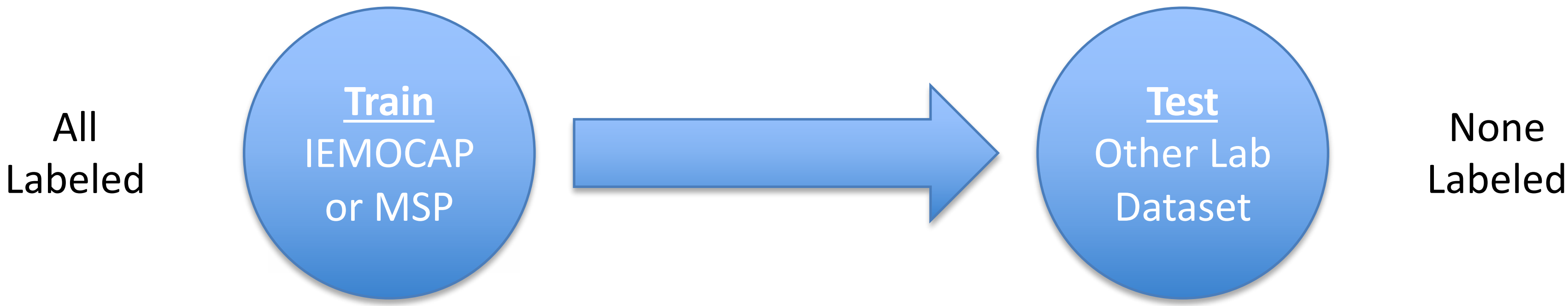
MADDoG: Multiclass ADDoG



Experimental Overview

- Four datasets:
  - IEMOCAP (16 kHz)
  - MSP-Improv (44.1 kHz)
  - PRIORI Emotion (8 kHz)
- Features: Mel Filterbanks (40d, length zero-padded to longest in batch)
- Task: cross-domain valence recognition (three-class)
- Setups:
  - Train on one lab dataset, test on another (IEMOCAP/MSP-Improv)
  - Train on one lab dataset, test on PRIORI Emotion
  - Train on two lab datasets, test on PRIORI Emotion

Experiment 1 – Cross Dataset

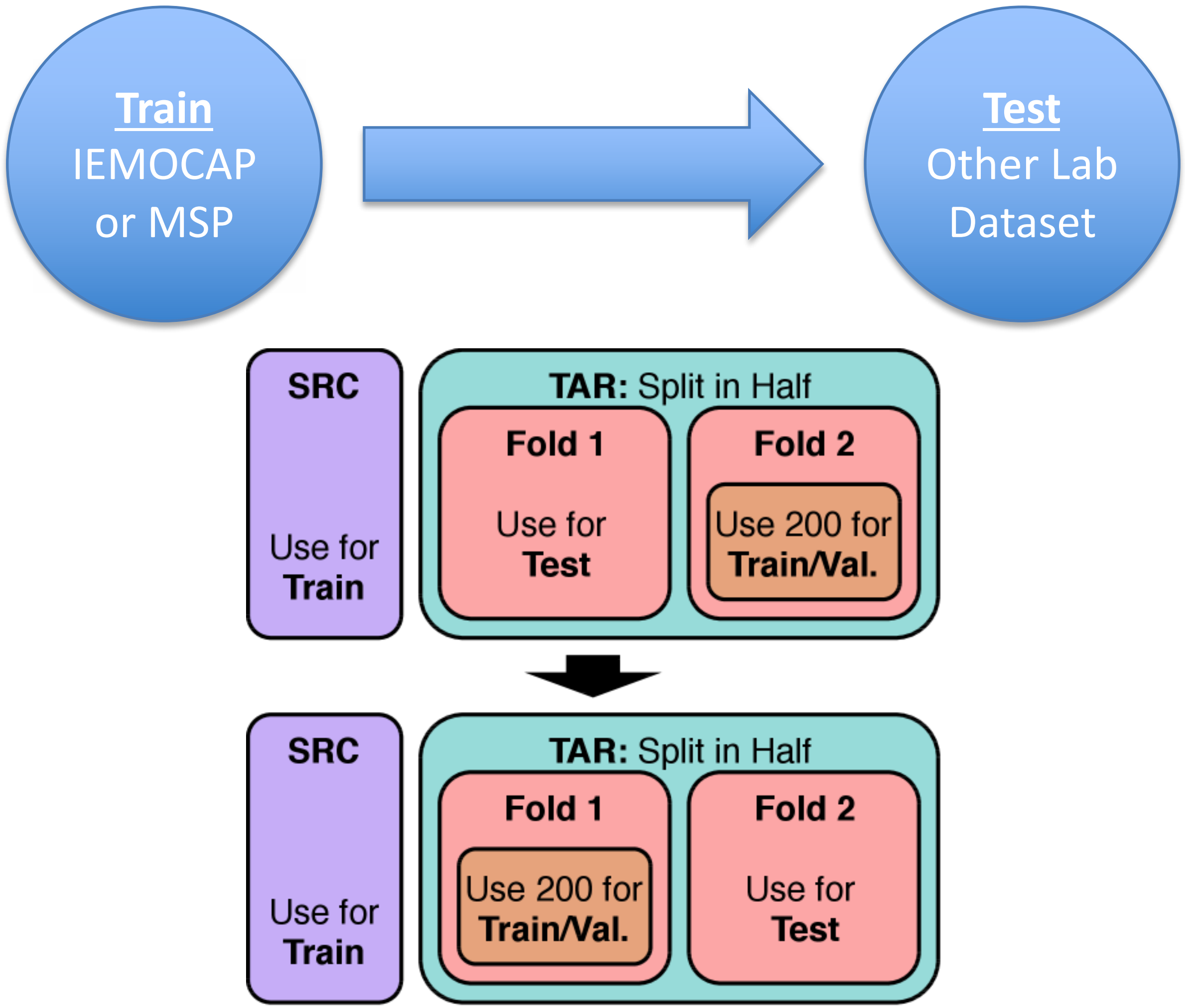


	MSP-Improv to IEMOCAP	IEMOCAP to MSP-Improv
CNN	0.439 ± 0.022 UAR	0.432 ± 0.012 UAR
ADDoG	**0.474 ± 0.009 UAR***	**0.444 ± 0.007 UAR***

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

Experiment 1 – Increasing Target Labels

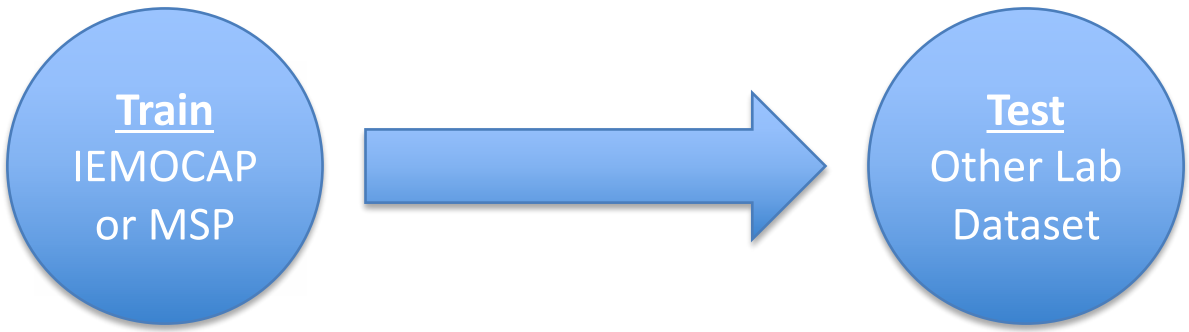
All Labeled



Some Labeled

Experiment 1 – Increasing Target Labels

All
Labeled



Some
Labeled

MSP-Improv to
IEMOCAP

IEMOCAP to
MSP-Improv

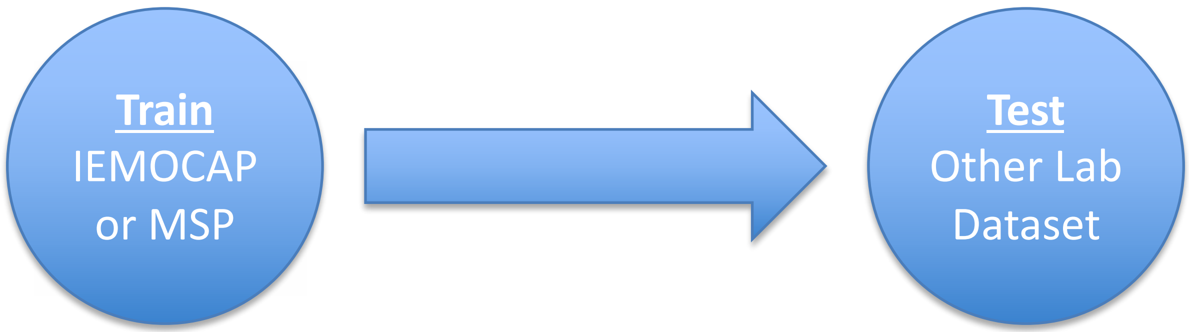
CHAI Lab
University of Michigan

Dots denote results significantly different than ADDoG (paired t-test, p=0.05)

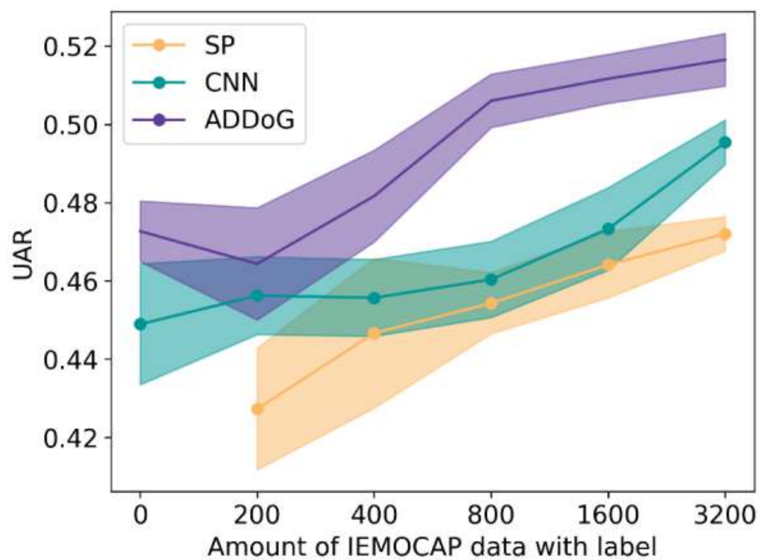
COMPUTER SCIENCE
& ENGINEERING
UNIVERSITY OF MICHIGAN

Experiment 1 – Increasing Target Labels

All Labeled



Some Labeled



MSP-Improv to IEMOCAP

IEMOCAP to MSP-Improv

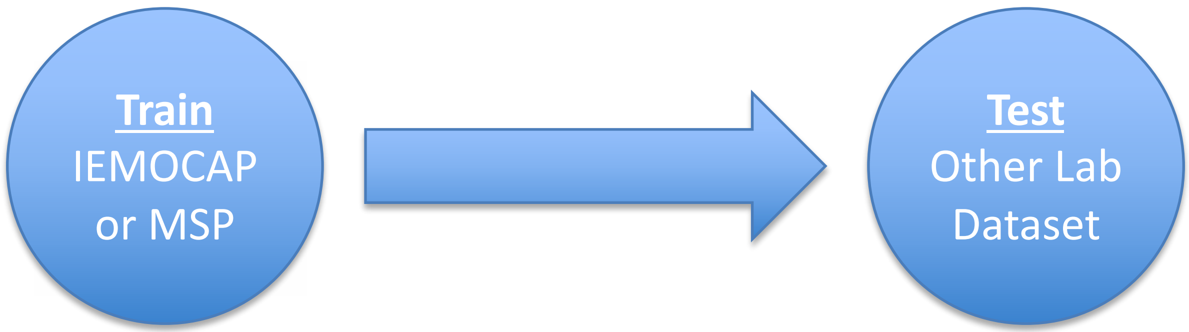
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Dots denote results significantly different than ADDoG (paired t-test, p=0.05)

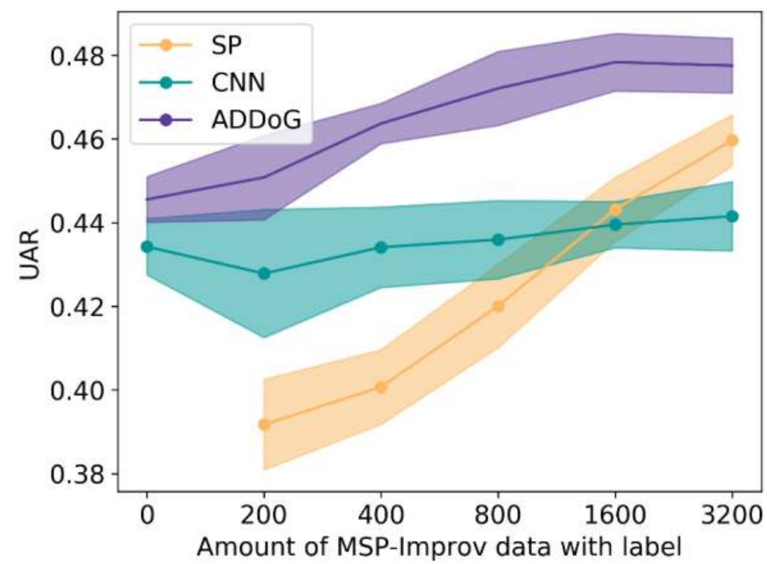
Experiment 1 – Increasing Target Labels

All Labeled



Some Labeled

MSP-Improv to IEMOCAP



IEMOCAP to MSP-Improv

CHAI Lab
University of Michigan

Dots denote results significantly different than ADDoG (paired t-test, p=0.05)

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Experiment 2 – To In-the-Wild Data

All Labeled

Train
IEMOCAP or MSP

Test
PRIORI Emotion

Some Labeled

IEMOCAP to PRIORI Emotion

MSP-Improv to PRIORI Emotion

CHAI Lab
University of Michigan

Dots denote results significantly different than ADDoG (paired t-test, p=0.05)

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

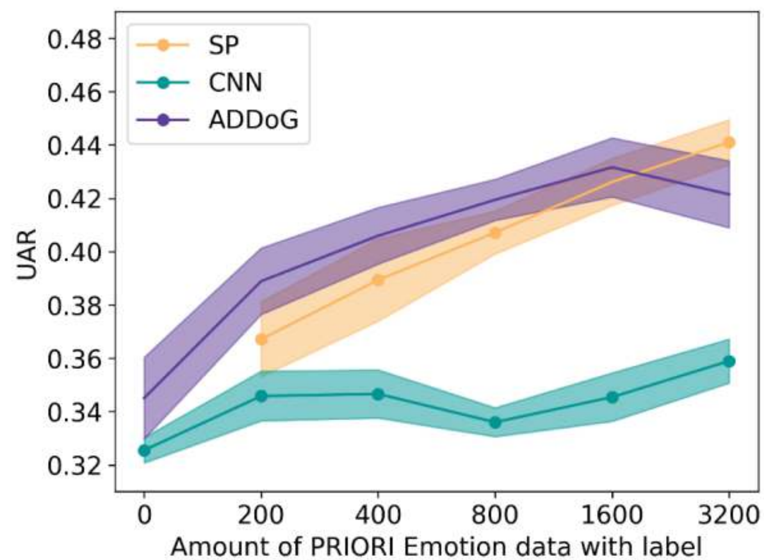
Experiment 2 – To In-the-Wild Data

All
Labeled

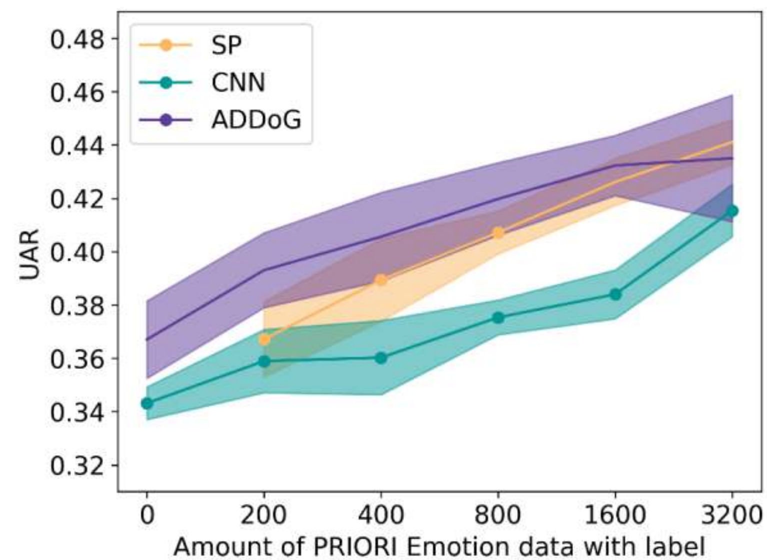
Train
IEMOCAP
or MSP

Test
PRIORI
Emotion

Some
Labeled



IEMOCAP to PRIORI Emotion



MSP-Improv to PRIORI Emotion

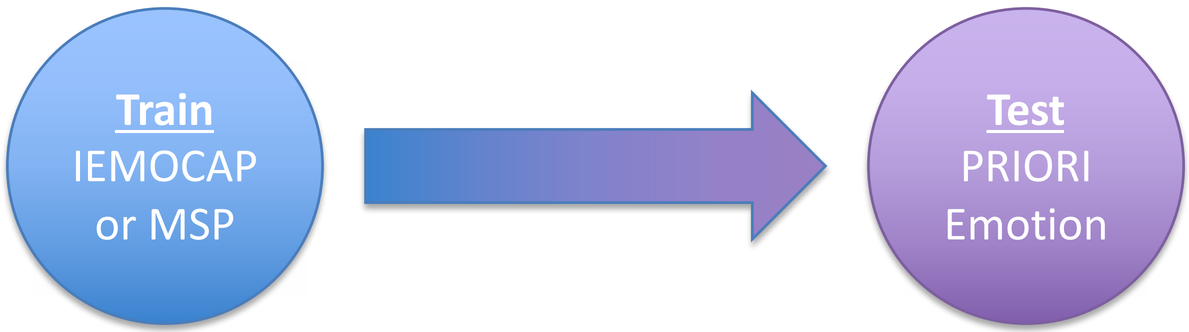
CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

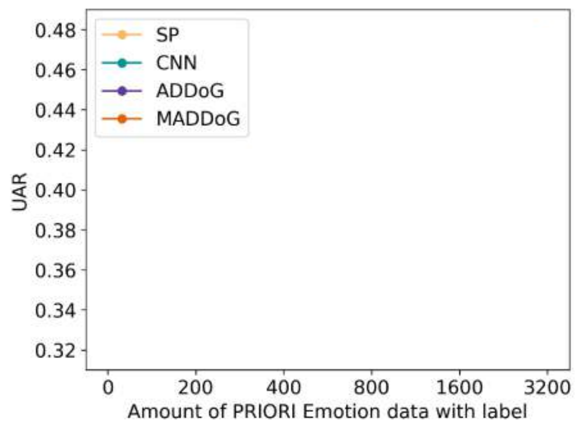
Dots denote results significantly different than ADDoG (paired t-test, p=0.05)

Experiment 3 – To In-the-Wild Data

All Labeled



Some Labeled



IEMOCAP and MSP-Improv to PRIORI Emotion

Dots denote results significantly different than MADDoG (paired t-test, p=0.05)

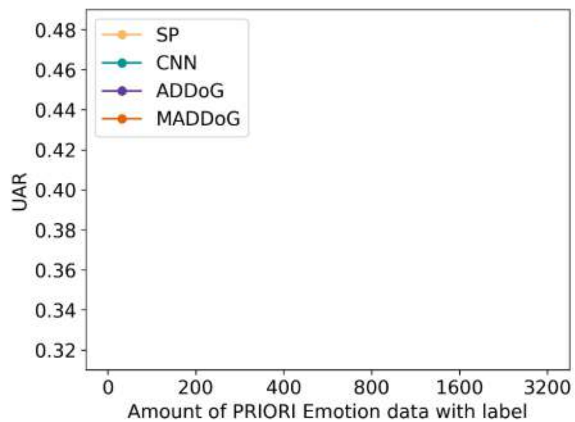
Experiment 3 – To In-the-Wild Data

All Labeled

Train
IEMOCAP and MSP

Test
PRIORI Emotion

Some Labeled



IEMOCAP and MSP-Improv to PRIORI Emotion

Dots denote results significantly different than MADDoG (paired t-test, p=0.05)

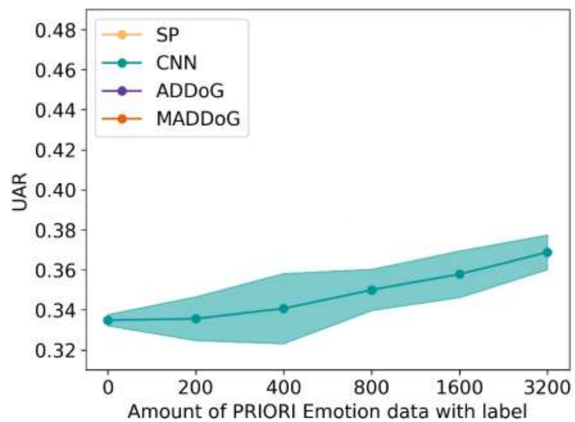
Experiment 3 – To In-the-Wild Data

All Labeled

Train
IEMOCAP and MSP

Test
PRIORI Emotion

Some Labeled



IEMOCAP and MSP-Improv to PRIORI Emotion

What we learn:
We *can't* train a model on outside datasets and expect them to just work

Dots denote results significantly different than MADDoG (paired t-test, p=0.05)

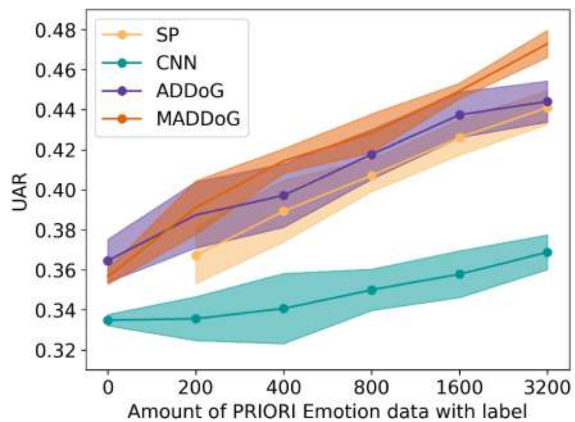
Experiment 3 – To In-the-Wild Data

All Labeled

Train
IEMOCAP and MSP

Test
PRIORI Emotion

Some Labeled



IEMOCAP and MSP-Improv to PRIORI Emotion

Dots denote results significantly different than MADDoG (paired t-test, p=0.05)

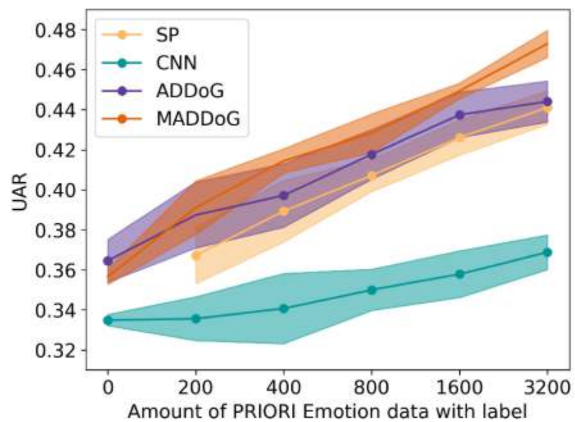
Experiment 3 – To In-the-Wild Data

All Labeled

Train
IEMOCAP and MSP

Test
PRIORI Emotion

Some Labeled



IEMOCAP and MSP-Improv to PRIORI Emotion

Where we can go:
We *can* use these models to derive emotion features in other domains [Interspeech 2019]

Dots denote results significantly different than MADDoG (paired t-test, p=0.05)

Conclusions

- ADDoG and MADDoG **consistently converge**
  - **Clear target** at each step (other dataset)
  - "**Meet in the middle**" approach
- Effective at detecting emotion in **smartphone calls**

CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

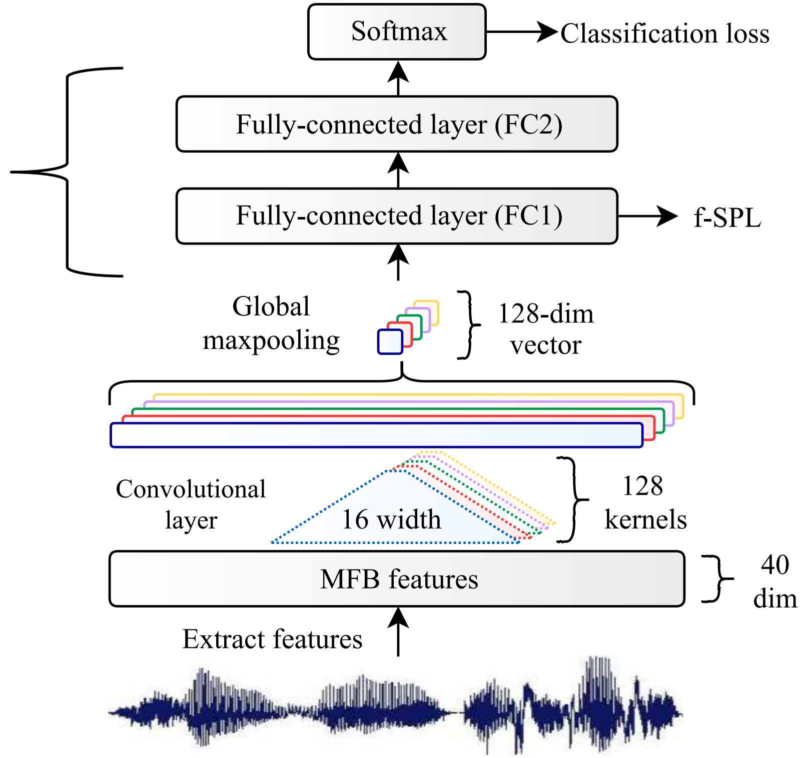
Remaining challenge:

We still aren't sure about the representation itself!

Emotion Recognition Representation

What if the representation held emotional meaning?

What if points close in emotion were close to each other?



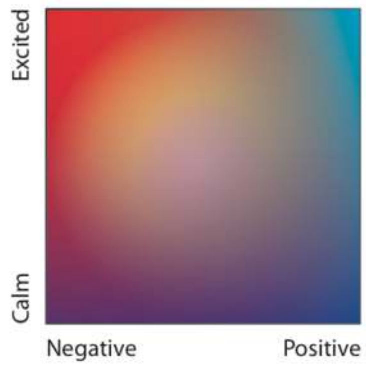
Biqiao Zhang, Yuqing Kong, Georg Essl, Emily Mower Provost. "f-Similarity Preservation Loss for Soft Labels: A Demonstration on Cross-Corpus Speech Emotion Recognition." AAAI. Hawaii. January 2019.

Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity

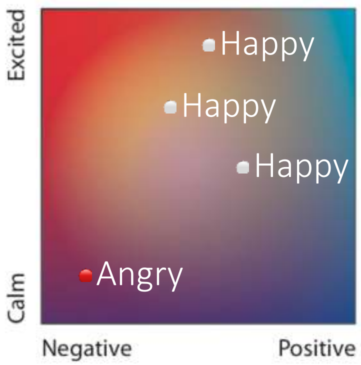
Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity



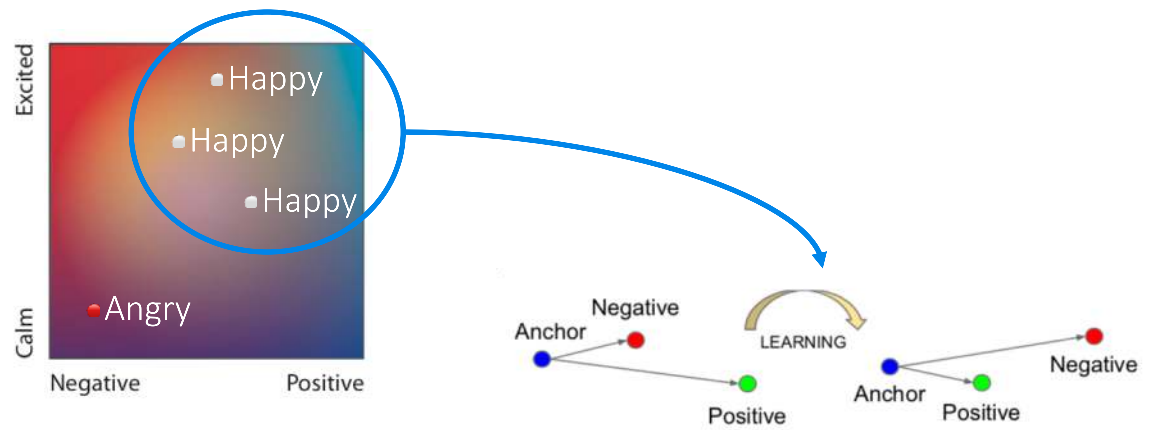
Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity



Deep Metric Learning (DML)

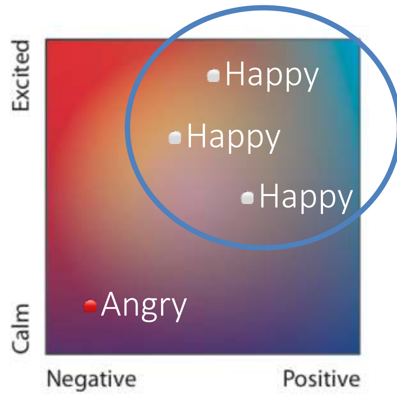
- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity

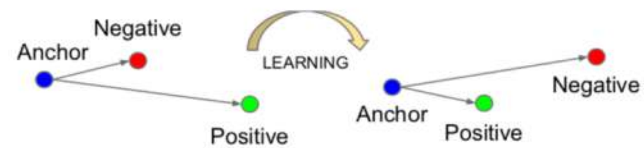


Triplet Loss
[Weinberger and Saul 2009; Chechik et al. 2010; Hoffer and Ailon 2015; Schroff et al. 2015]

Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity

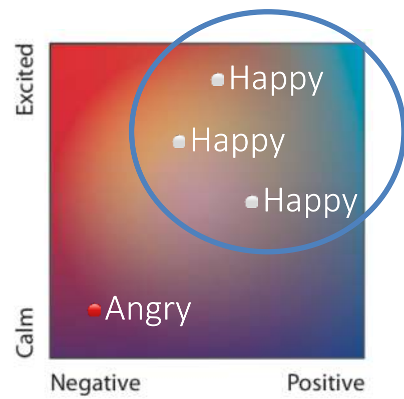


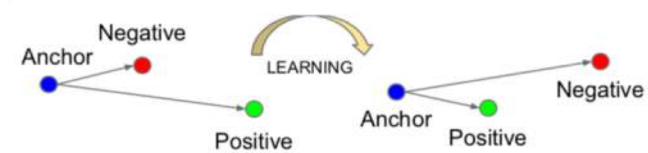


Triplet Loss
[Weinberger and Saul 2009; Chechik et al. 2010; Hoffer and Ailon 2015; Schroff et al. 2015]

Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity

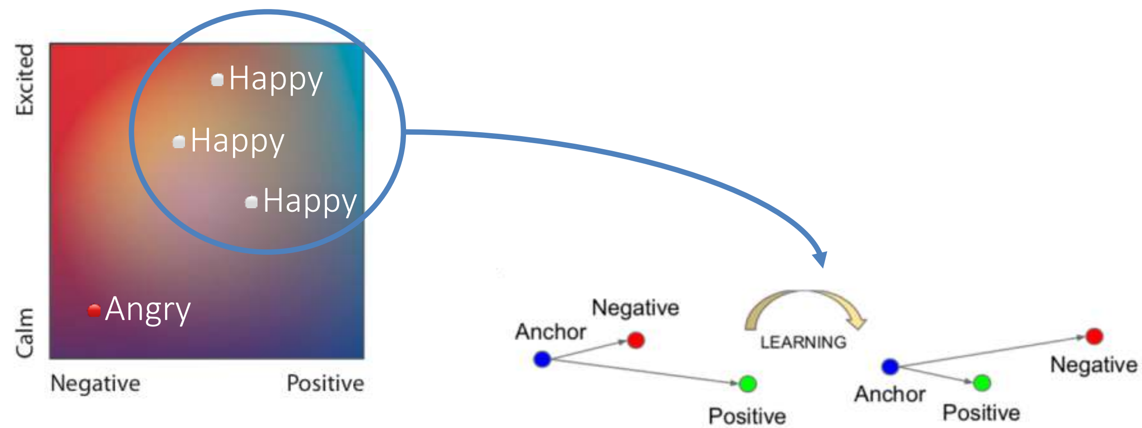




Triplet Loss
[Weinberger and Saul 2009; Chechik et al. 2010; Hoffer and Ailon 2015; Schroff et al. 2015]

Deep Metric Learning (DML)

- **Goal**: learn an embedding space where pairwise distance corresponds to label similarity

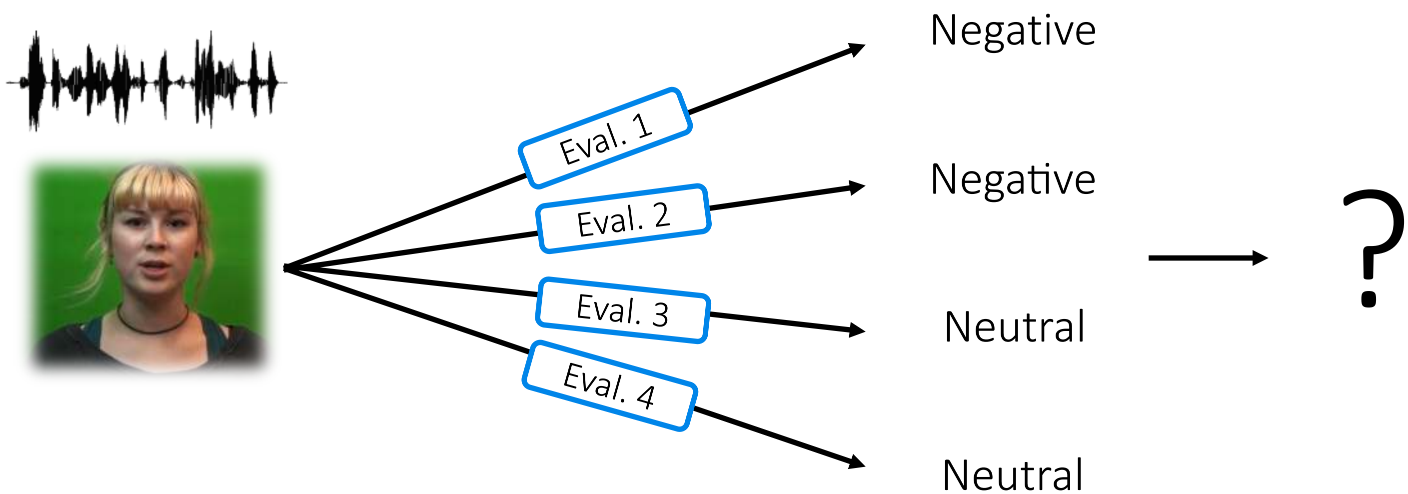


Triplet Loss
[Weinberger and Saul 2009; Chechik et al. 2010; Hoffer and Ailon 2015; Schroff et al. 2015]

Variability is signal, not just noise

Hard labels are too limiting.

- Disagreement in evaluation is extremely common

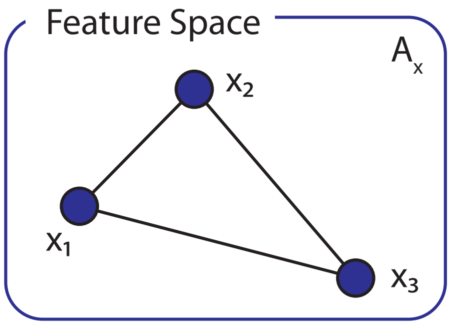


f-Similarity Preservation Loss

Goal: learn an embedding space where feature similarity = label similarity

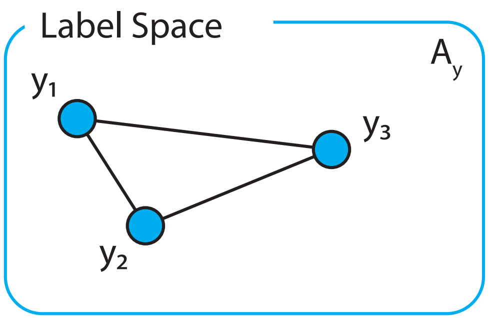
f-Similarity Preservation Loss

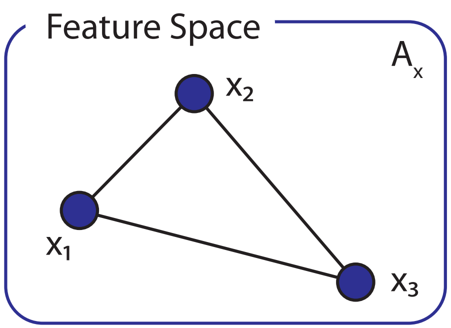
Goal: learn an embedding space where feature similarity = label similarity



f-Similarity Preservation Loss

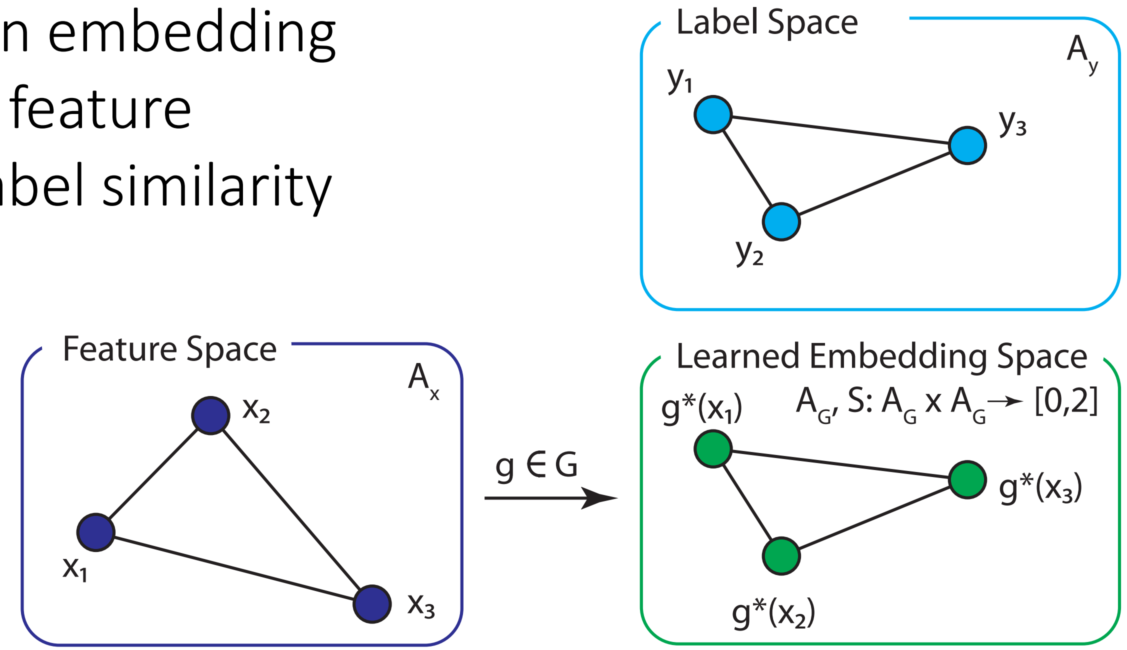
Goal: learn an embedding space where feature similarity = label similarity





f-Similarity Preservation Loss

Goal: learn an embedding space where feature similarity = label similarity

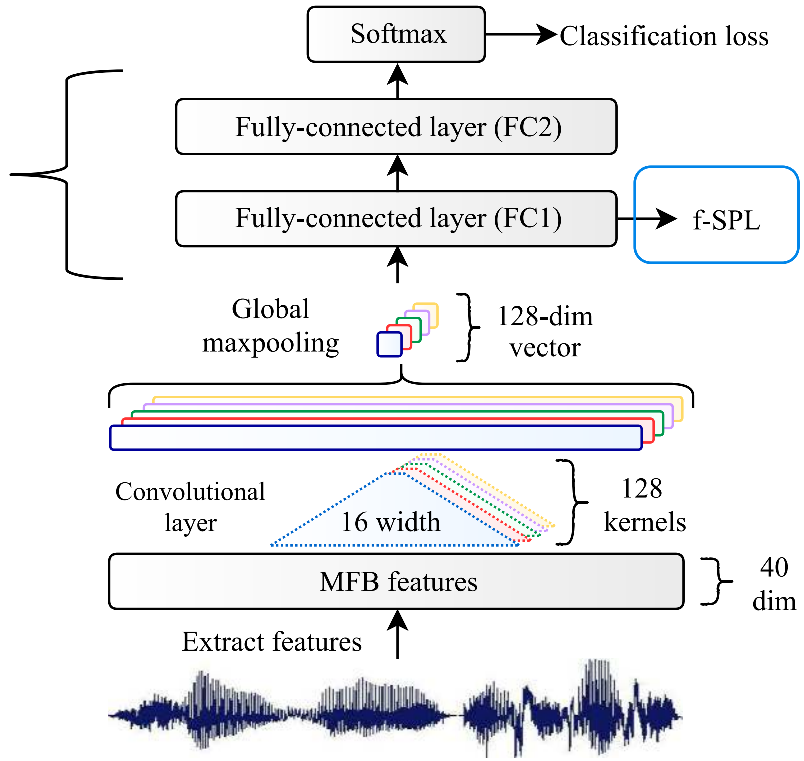


CHAI Lab
University of Michigan

COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

New Representations

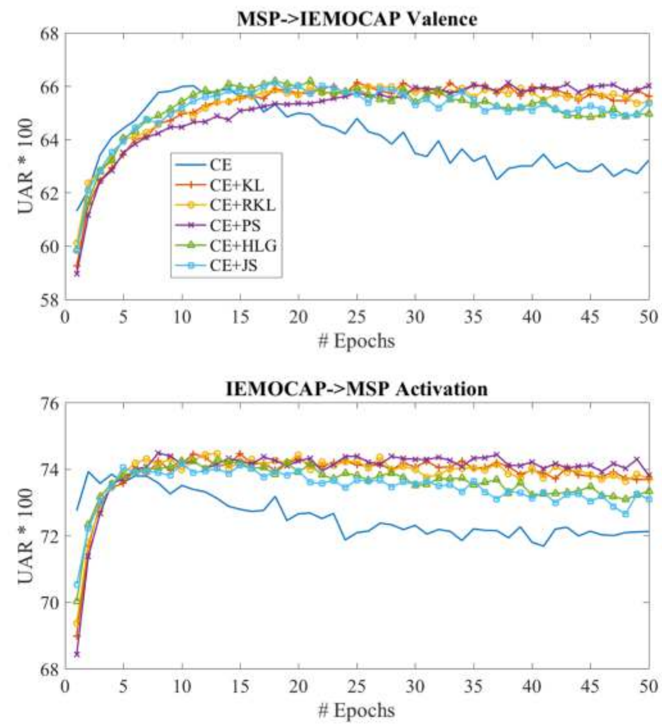
Enforce emotional meaning!



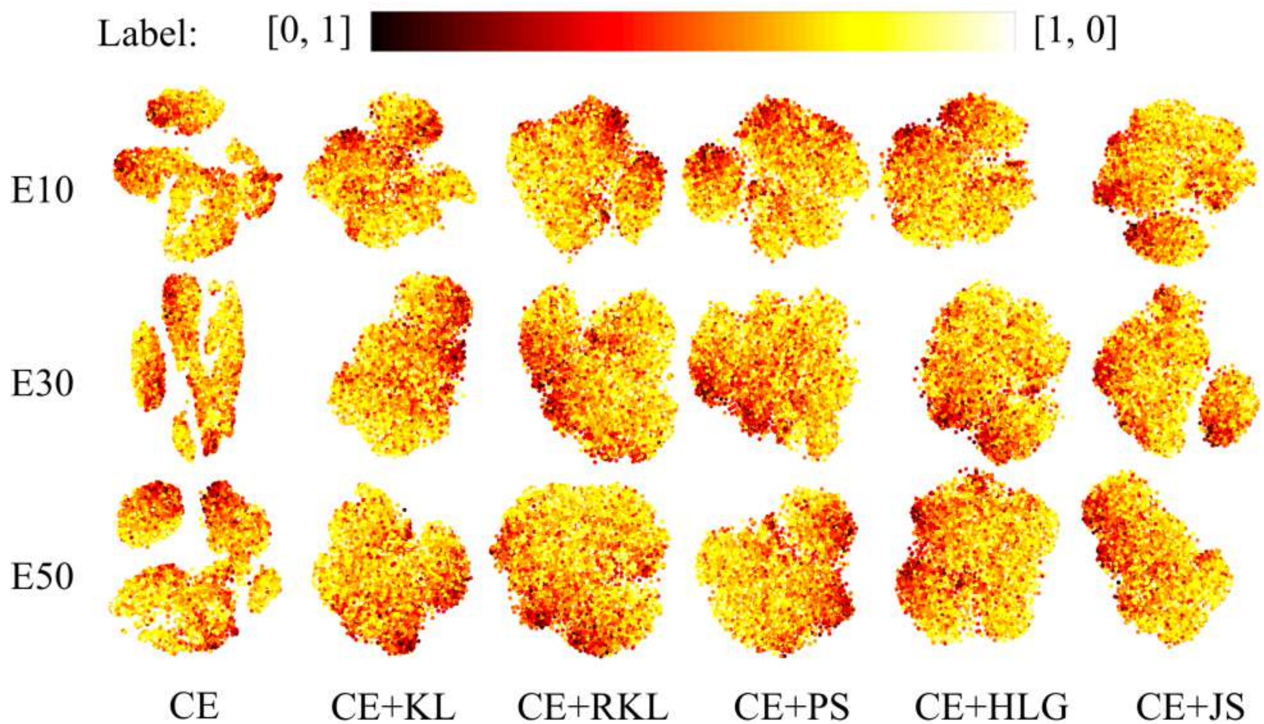
Performance on heldout data

- *f*-SPL less susceptible to overfitting

- Statistically significantly higher performance compared to cross-entropy loss

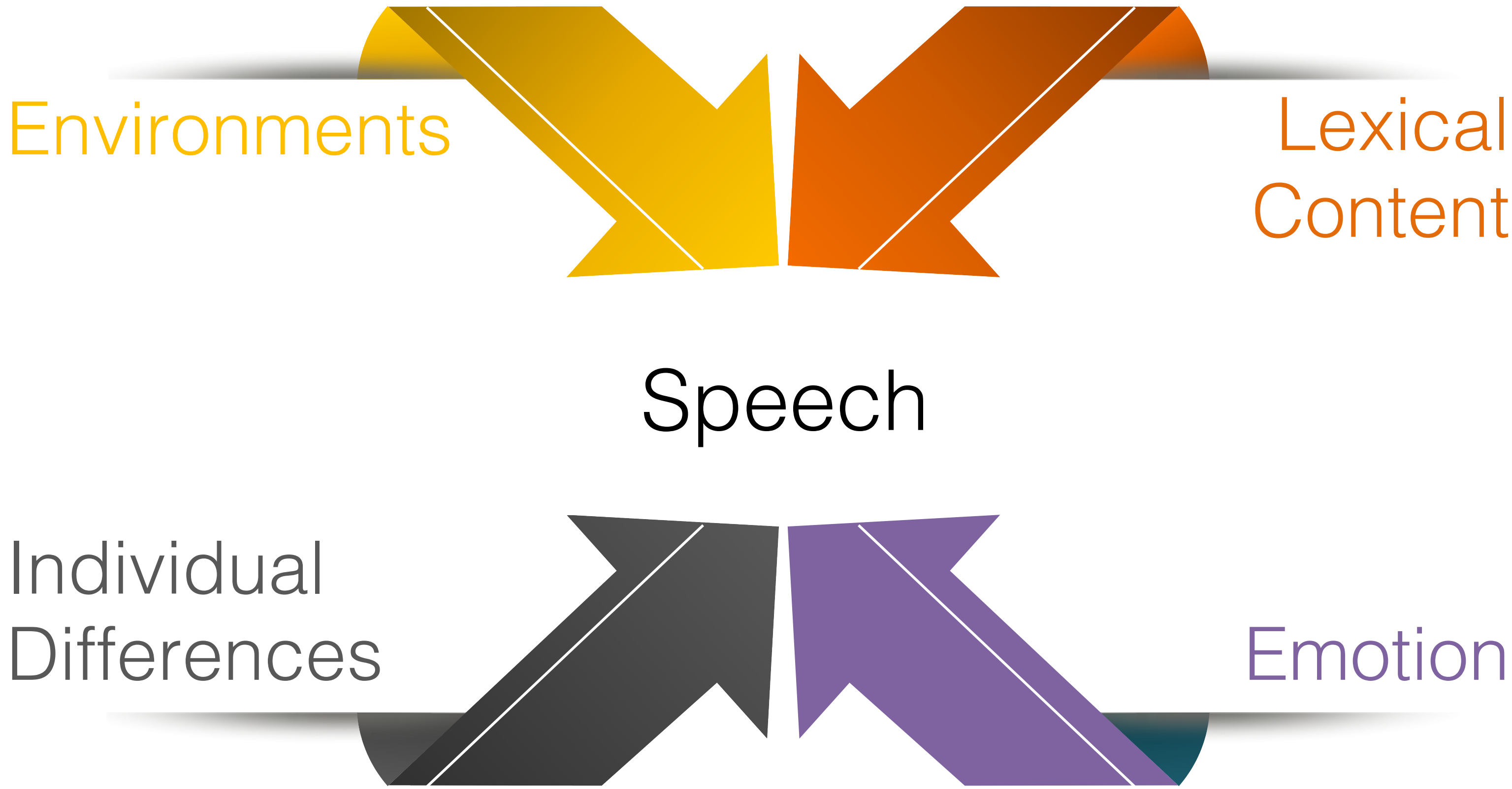


Embedding with emotional meaning



Baseline

Embracing Complexity



Thanks!

Questions?

TOYOTA RESEARCH INSTITUTE

IBM Watson

NATIONAL INSTITUTES OF HEALTH

NSF

MIDAS

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

University of Michigan Depression Center

CHAI Lab University of Michigan

COMPUTER SCIENCE & ENGINEERING UNIVERSITY OF MICHIGAN