USC University of Southern California

Human-centered Multimodal Machine Intelligence

Shrikanth (Shri) Narayanan
Signal Analysis and Interpretation Laboratory (SAIL)
http://sail.usc.edu

University of Southern California

February 9, 2022
University of Miami
IDSC Lecture Series

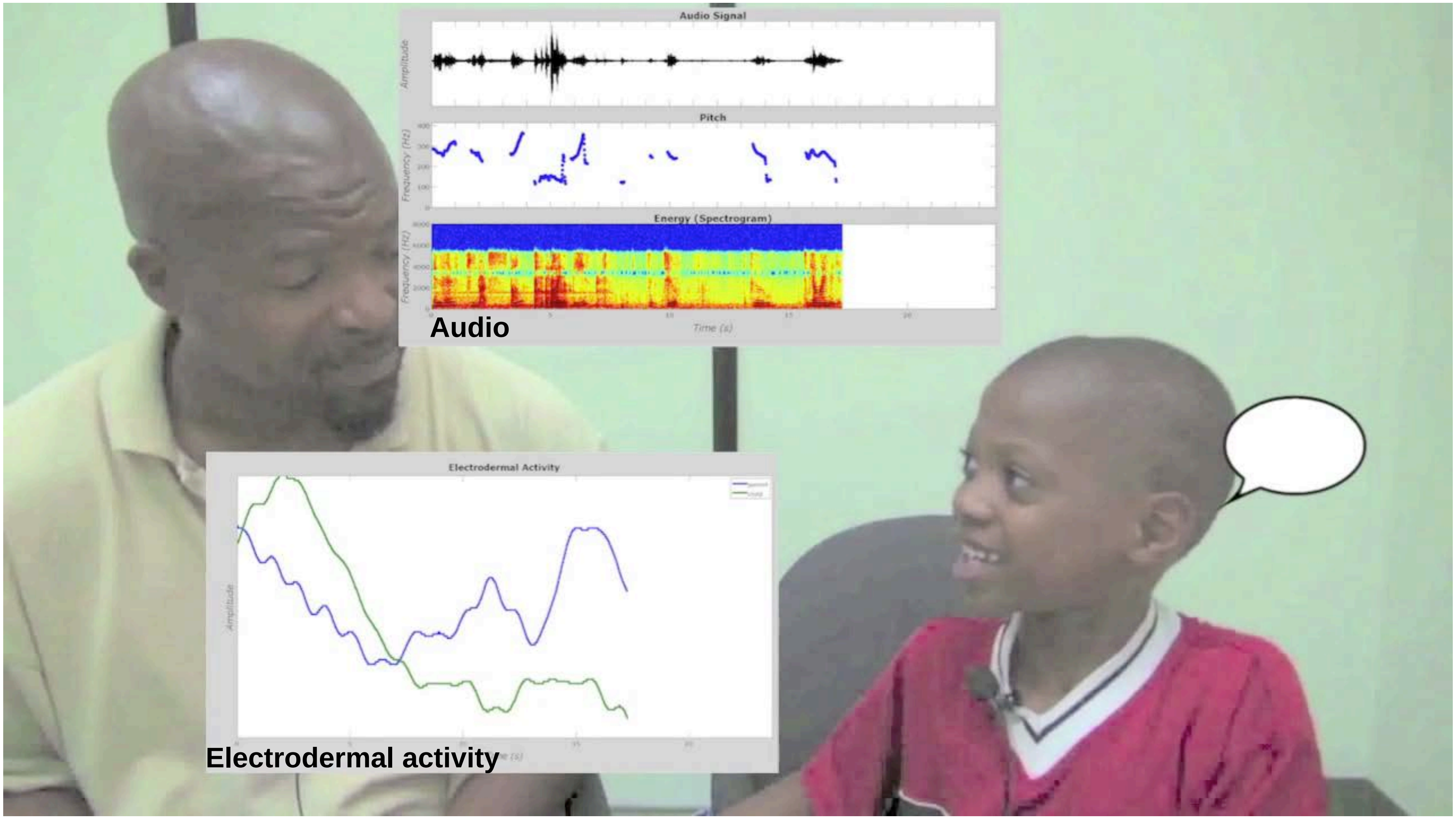
Human-centered Machine Intelligence: Promise & Possibilities

- **Exciting converging advances**
  - *technologies*: sensing, computing, machine learning, data communication, interfaces (e.g., devices on/with/by people)
  - *people*: amazing cross-disciplinary partnerships, resource sharing across societal application domains

Human-centered Machine Intelligence: Promise & Possibilities

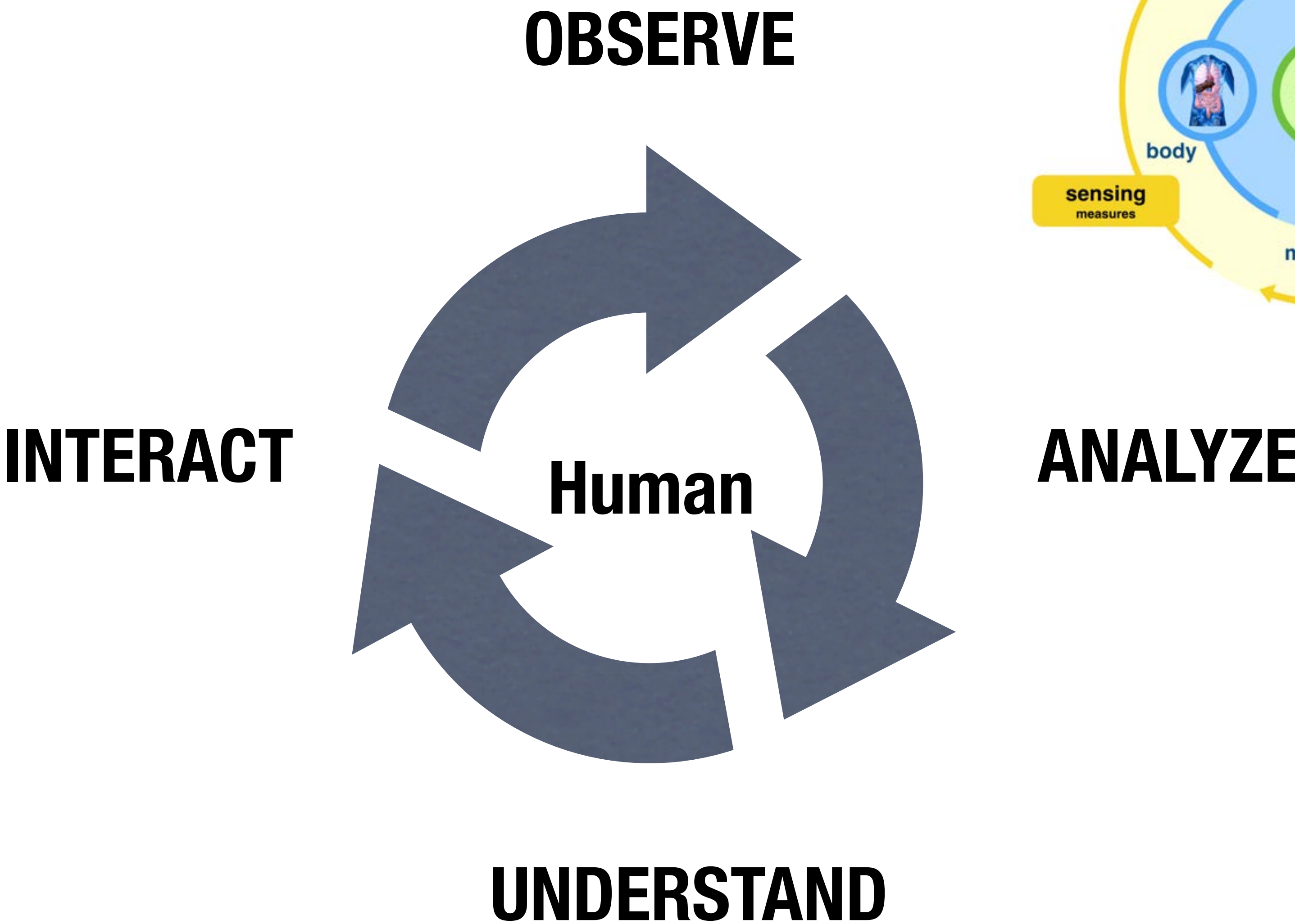
- **Exciting converging advances**
  - *technologies*: sensing, computing, machine learning, data communication, interfaces (e.g., devices on/with/by people)
  - *people*: amazing cross-disciplinary partnerships, resource sharing across societal application domains
- **Novel possibilities to help understand, support, and enhance the human experience**
- **Challenge and Opportunity: Technologies that**
  - work for *everyone* and in *all* contexts
  - understand and support the rich variety in discerning the *who, what, where, how, when*,…
    - *Speech and spoken language offer a key source of information*
  - *Example*: machine intelligence to analyze human conversation

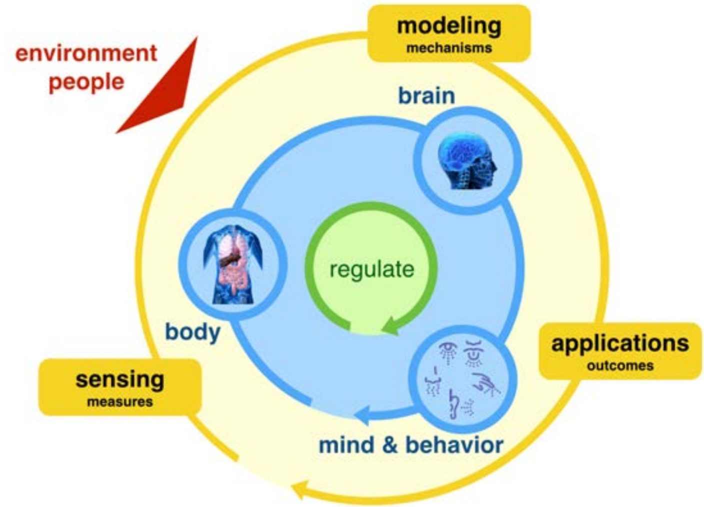
Rich Understanding of Multimodal Behavior and Interaction



Example: Parent and child creating a story together

HUMAN-CENTERED MACHINE INTELLIGENCE ECOSYSTEM

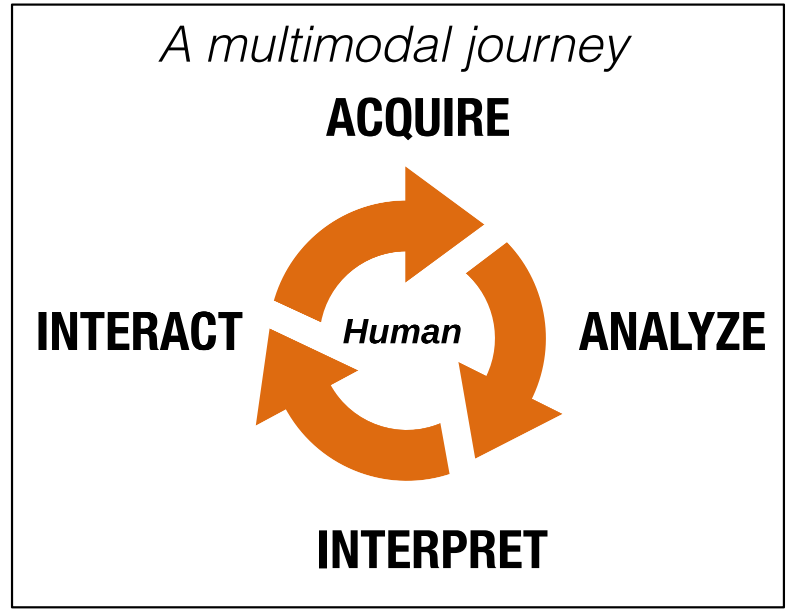


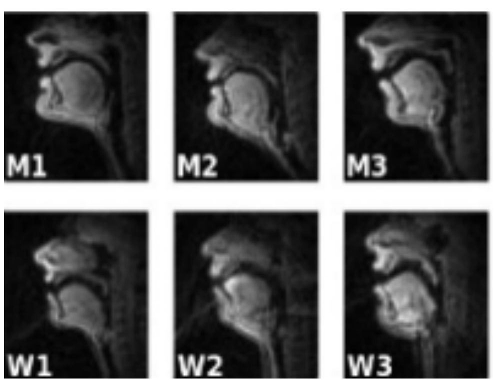


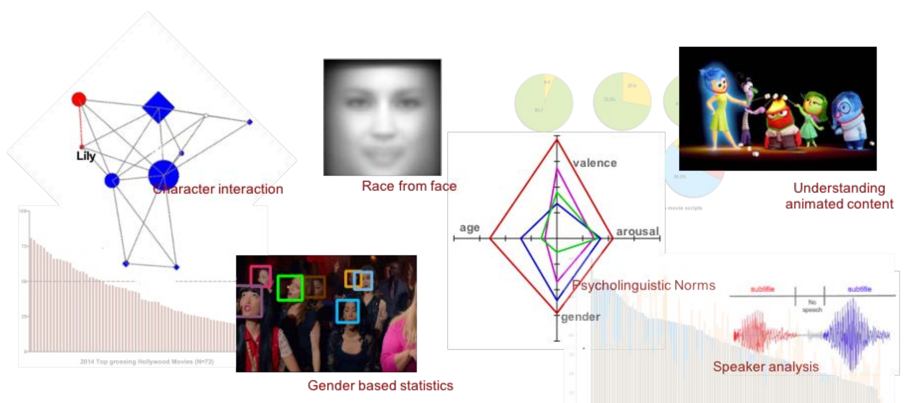
Shrikanth S. Narayanan and Asad M Madni. Inclusive Human centered Machine Intelligence. The Bridge, 50:113–116, National Academy of Engineering, 2020.

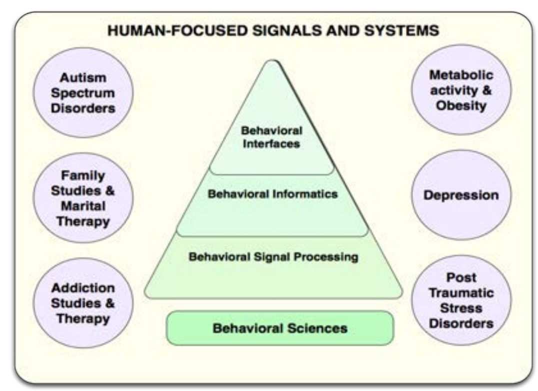
Highlights from two areas of Human-centered Machine Intelligence

✓ Behavioral Machine Intelligence
✓ Computational Media Intelligence









Highlight 1

Behavioral Machine Intelligence

Psychological Health and Well being applications

From Wearable & Environmental Sensing to Artificial Intelligence Methods

- engineering approaches to illuminate human trait and mental state
- screening, diagnostic, intervention support in mental and behavioral health

SUPPORT FROM NIH, NSF, DoD, IARPA, SIMONS FOUNDATION

Seeking a window into human trait, state and behavior

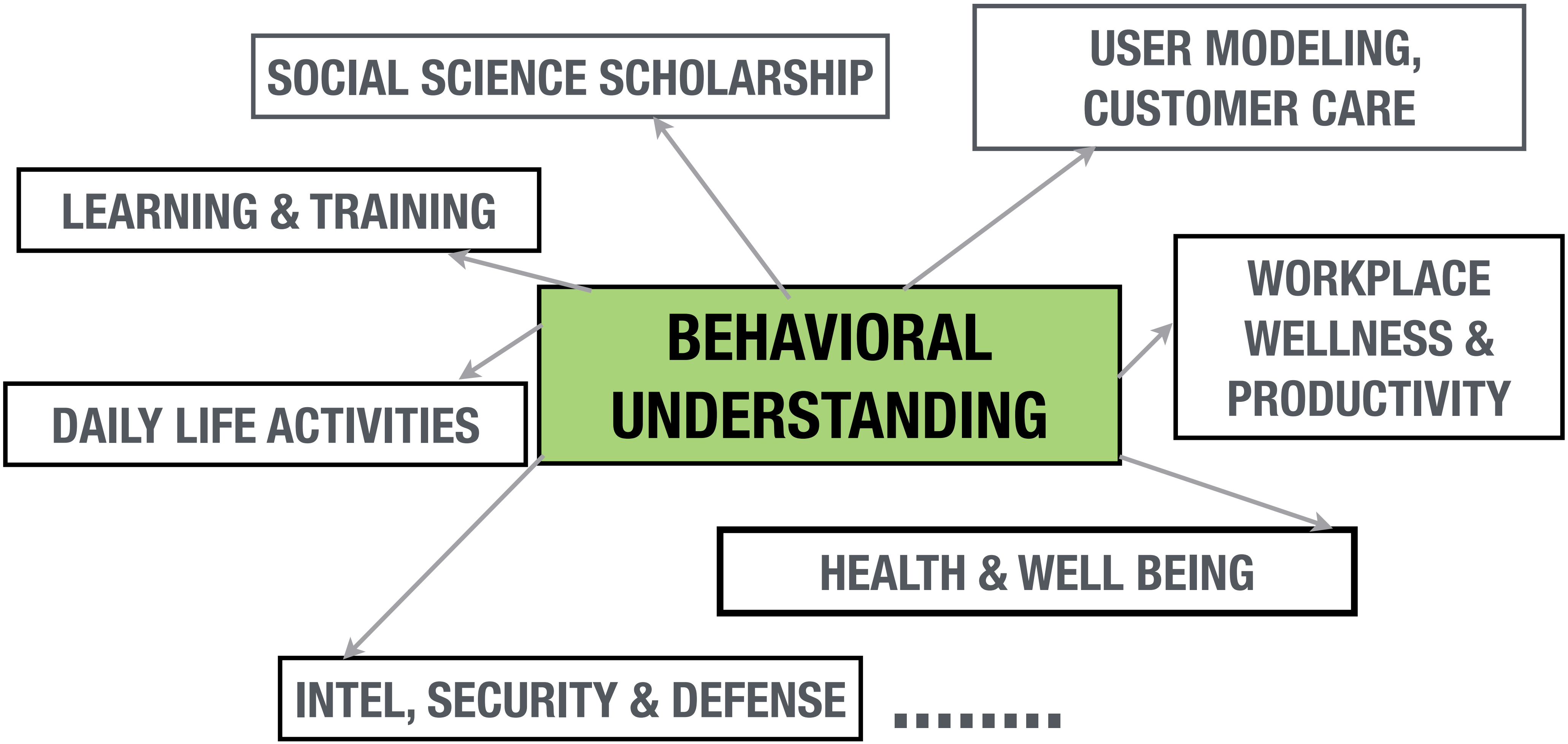
using engineering approaches and technologies

.....from qualitative to quantitative

Scalable, Broadly Accessible, Cost Effective

UNDERSTANDING BEHAVIOR CENTRAL TO MANY HUMAN DOMAINS

... ACROSS APPLICATIONS

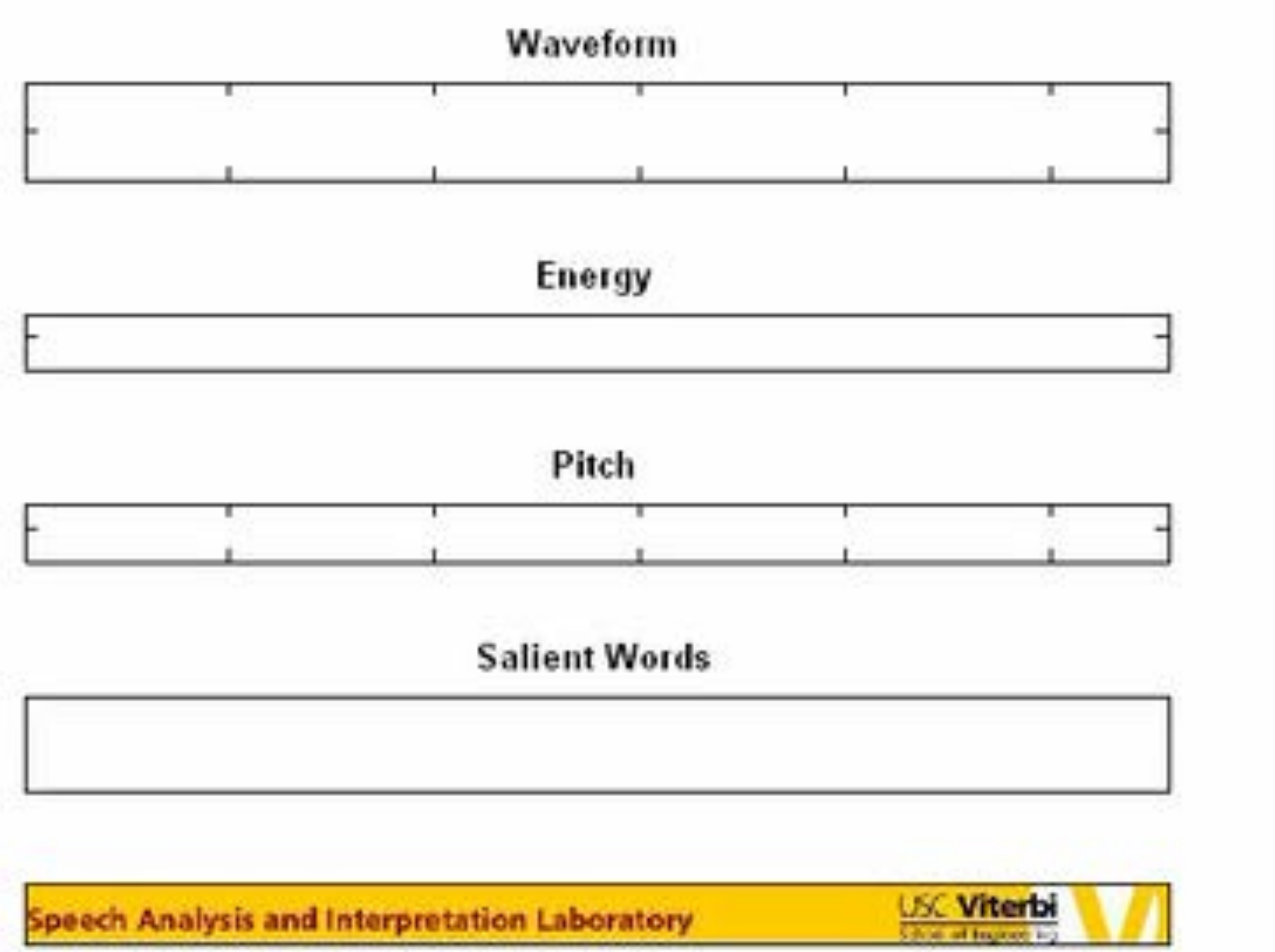


ROLE OF ENGINEERING?

Customer care

circa 2001

Escalating frustration? **(only customer side played)**



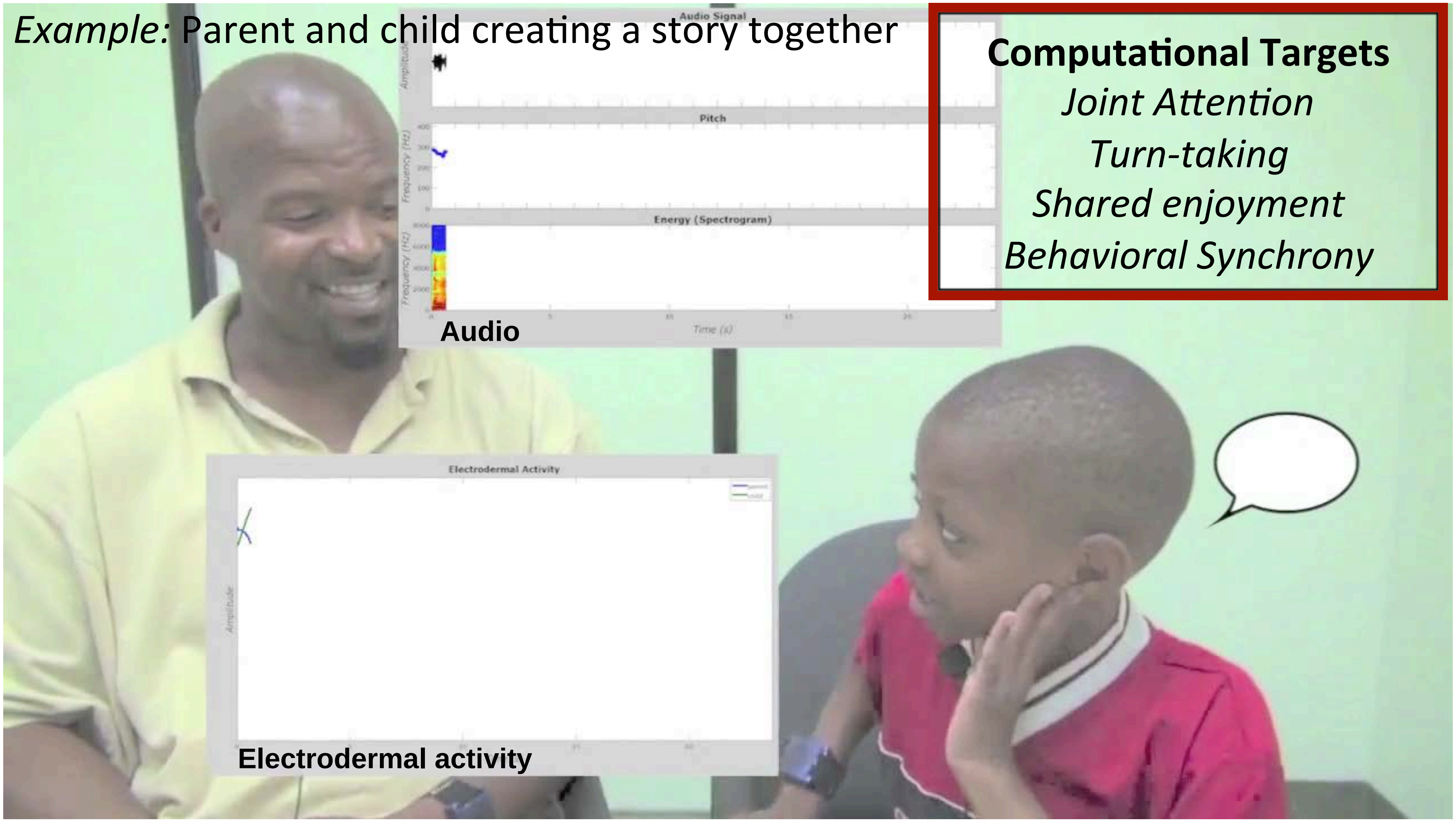
PREVALENCE OF SELECTED HEALTH CONDITIONS (IN THE US)

Condition	Ages	Prevalence*
Autism spectrum disorder	Children (typically diagnosed as children, but persist over lifetime)	1.5% (lifetime)
Posttraumatic stress disorder	Adults	3.5% (one year)
Mood disorders (e.g., depression)	Adults	9.5% (one year)
Alcohol addiction/abuse	All	6.6% (one year)
Illicit drug use (nonmarijuana)	All	2.5% (one year)
Parkinson's disease	> 60 years old	1.9% (lifetime)
Dementia (e.g., Alzheimer's disease)	> 65 years old	6.5% (lifetime)

***Sources listed in:**
Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications. IEEE Signal Processing Magazine. 34(5): 189-196, September 2017

Autism Spectrum Disorder

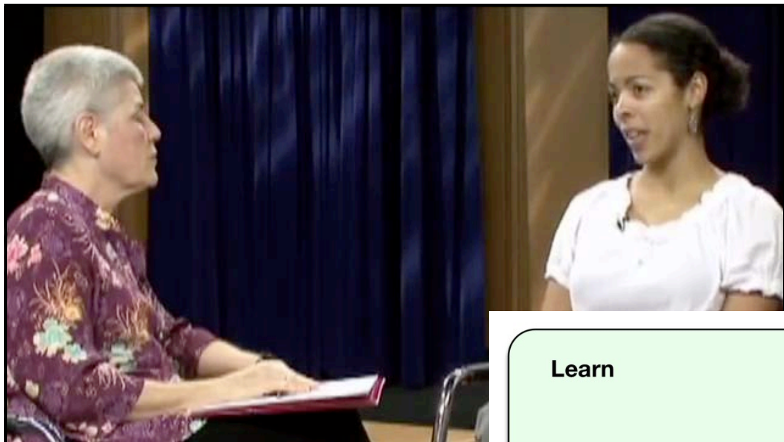
Technologies for Rich Understanding of Expressive Behavior and Interaction?



- **1 in 54 US children diagnosed with ASD (CDC, 2020)**
- **ASD characterized by difficulties in social communication, reciprocity; repetitive or stereotyped behaviors and interests**

Addiction treatment: Psychotherapy

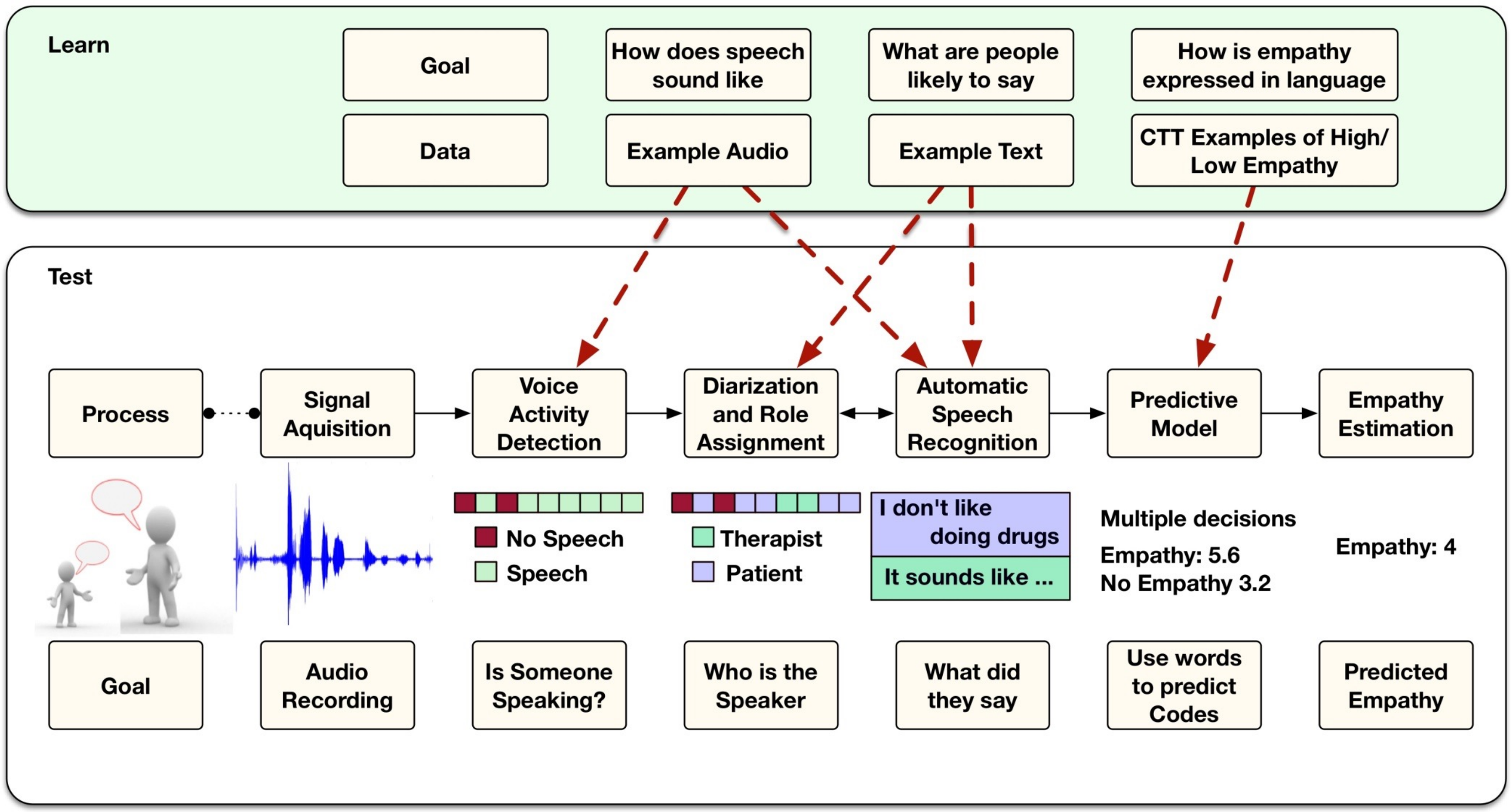
Illuminating what works, for whom, how and why



Motivational Interviewing

https://www.youtube.com/watch?v=EvLquWI8aqc

Empathy expressed by the therapist?

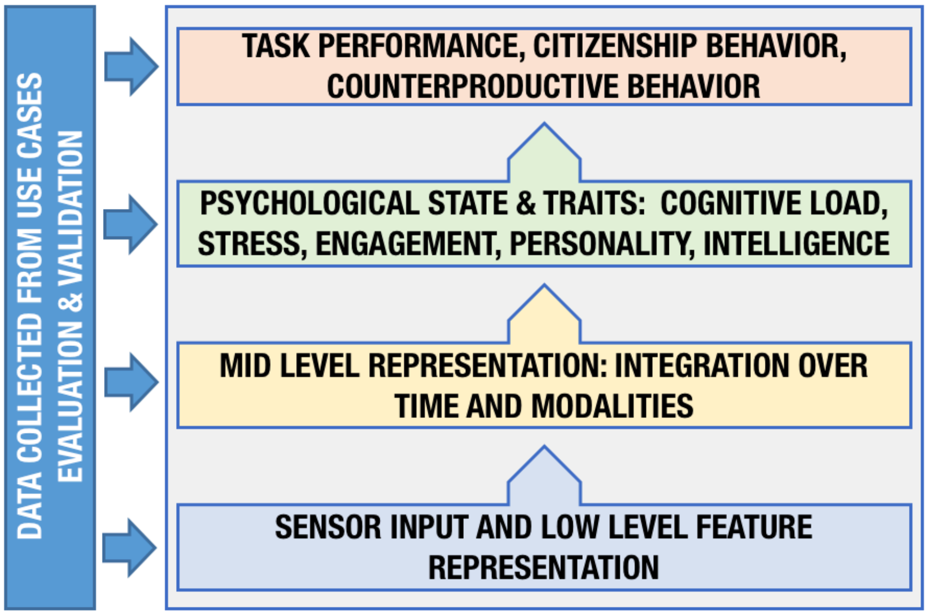


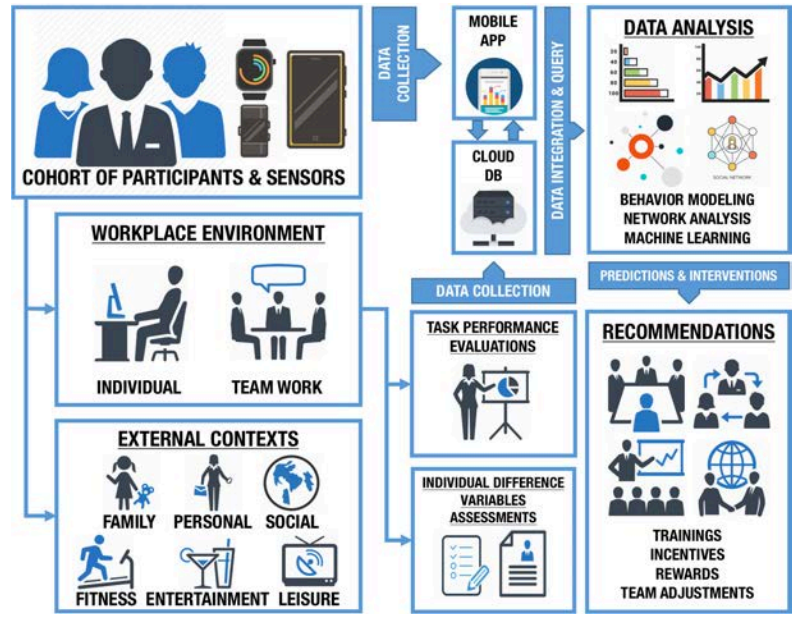
Annual costs of addiction exceed $740 Billion

https://www.drugabuse.gov/related-topics/trends-statistics

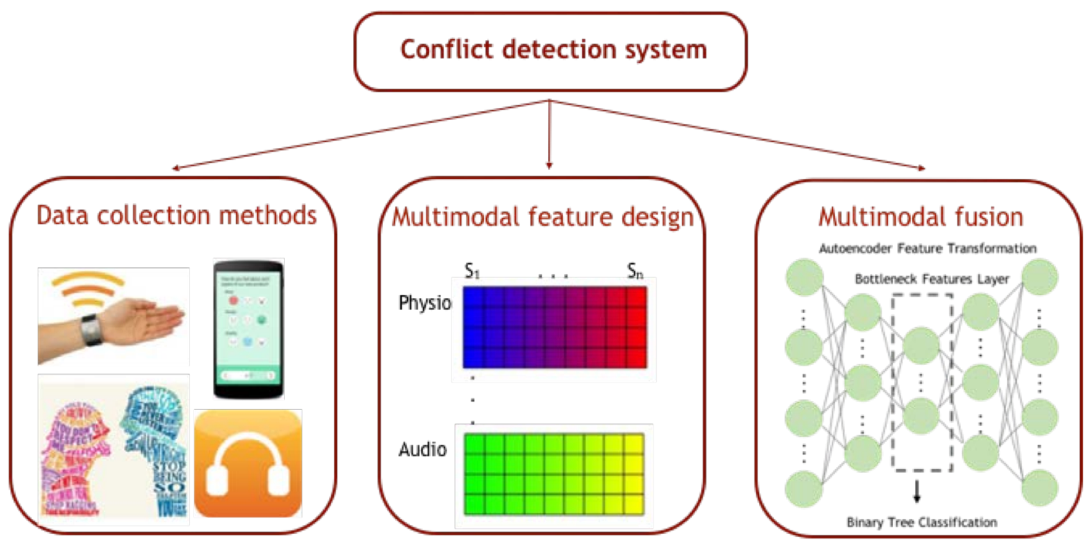
Day-to-day Health, Well being: Home, Work place

Bio-behavioral & IoT platform for individualized assessment and support





Stress? Conflict?



HUMAN-CENTERED MACHINE INTELLIGENCE:

SUPPORT DECISION MAKING, ACTION & RESPONSE USING SENSING, DATA SCIENCES AND AI TECHNOLOGIES

✓ **HELP US DO THINGS WE KNOW TO DO MORE EFFICIENTLY, CONSISTENTLY**

✓ **HELP HANDLE NEW DATA, CREATE NEW MODELS TO OFFER NOVEL INSIGHTS**

- ✓ **CREATE TOOLS FOR SCIENTIFIC DISCOVERY**

✓ **HELP CREATE TOOLS TO SUPPORT DIAGNOSTICS, PERSONALIZED INTERVENTION, AND TRACKING ITS RESPONSE TO TREATMENT**

Multimodal Bio-Behavior & Context Signals

- **"Sounds, Words, Sight" offer a peek into traits and (hidden) human state**
  - speech, language use, dialogic interaction,
  - accompanied by facial expressions, body language,…
  - and, perhaps complemented by physiological (e.g., ECG) and neural measures (e.g., EEG)
- **Environmental measures for context**

  e.g., location, temperature, light, sound, humidity, air qlty,..

MEASURE & QUANTIFY HUMAN BEHAVIOR

CONFLUENCE OF SENSING, COMMUNICATION AND COMPUTING

Operationalizing…

Behavioral Machine Intelligence

- ***nuts and bolts***: foundational multimodal signal processing of data
  - *from people*: audio/speech, video, text, biosignals (ECG, EEG),..
  - *from the environment*: location, temperature, light, sound, humidity, air quality,..
- ***construct prediction***: machine learning based methods for automated behavioral coding and characterization
- ***computational modeling***: of interaction processes & mechanisms
- ***translational applications notably in health***: screening, diagnostics, intervention support
  - JIT implementation, tracking response to treatment,..

SHIFT TO MODELING MORE ABSTRACT, DOMAIN-RELEVANT CONSTRUCTS
……NEEDS NEW MULTIMODAL COMPUTATIONAL APPROACHES

Shrikanth Narayanan and Panayiotis Georgiou. Behavioral Signal Processing: Deriving Human Behavioral Informatics from Speech and Language. Proceedings of IEEE. 101(5): 1203-1233, May 2013

Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications. IEEE Signal Processing Magazine. 34(5): 189-196, September 2017

How is technology helping already? deep or not

- **Significant advances in foundational aspects of behavior modeling: detect, classify and track**
  - Audio & Video diarization: who spoke when; doing what,..
  - Speech recognition: what was spoken
  - Visual activity recognition: head pose; face/hand gestures
  - Physiological signal processing with EKG, GSR, ..
  - IoT technologies for environmental modeling and edge processing

SIGNAL PROCESSING AND MACHINE LEARNING ARE KEY ENABLERS

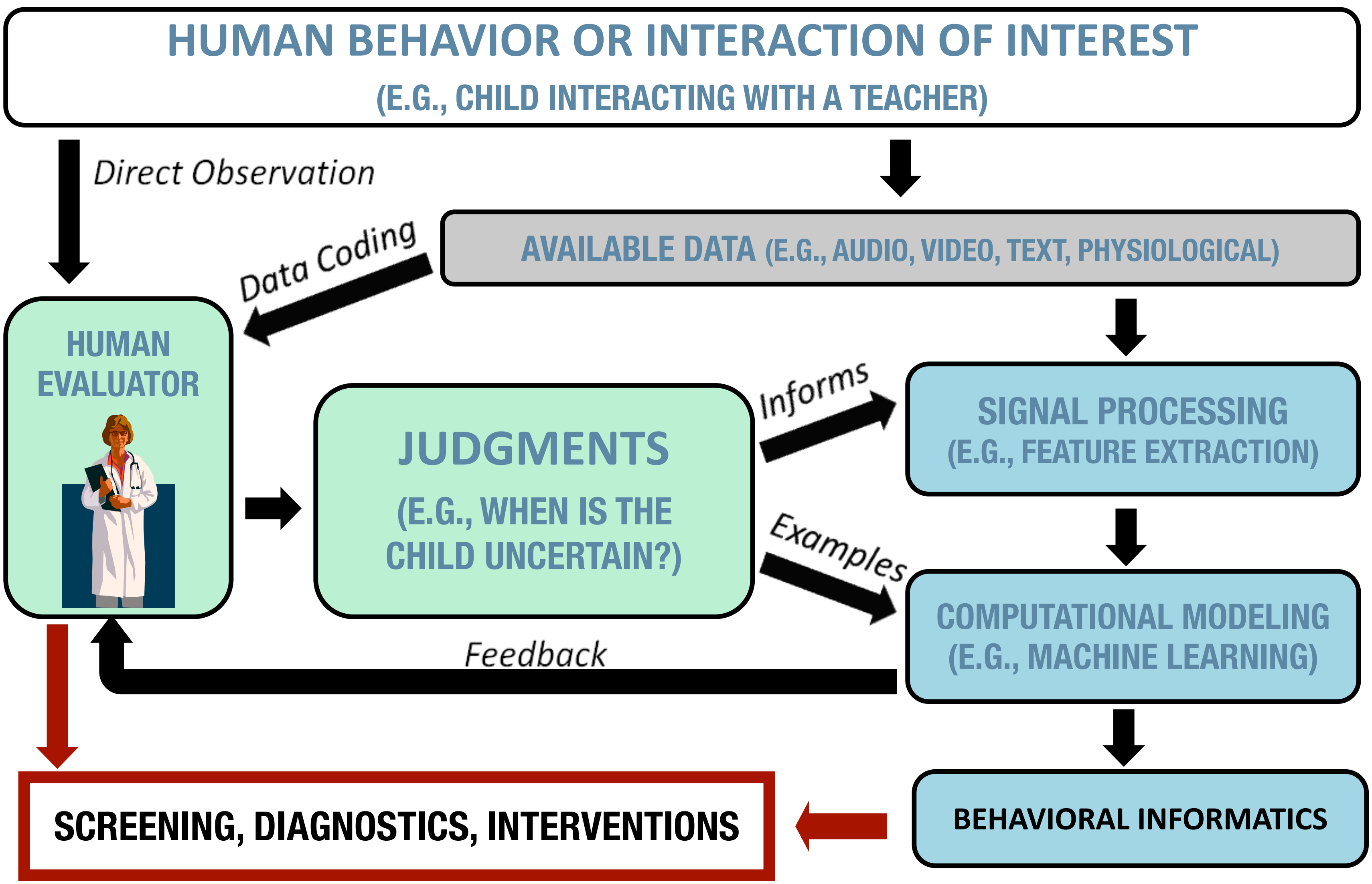
Example: A range of speech/language technology possibilities

- VOICE ACTIVITY DETECTION
- AUDIO SEGMENTATION
- ALIGNMENT
- TRANSCRIPTION
- KEYWORD SPOTTING
- PROSODY MODELING: INTONATION, PHRASING, PROMINENCE
- VOICE QUALITY
- NATURAL LANGUAGE PROCESSING OF TEXT/TRANSCRIPTS
- DIALOG ACT TAGGING
- INTERACTION MODELING: TURN TAKING DYNAMICS, ENTRAINMENT
- SPEAKER/VERIFICATION IDENTIFICATION
- AFFECTIVE COMPUTING FROM SPEECH AND LANGUAGE
- SPEAKER STATE AND TRAIT CHARACTERIZATION
- JOINT SPEECH AND VISUAL CUE PROCESSING

WITH VARYING DEGREES OF TECHNOLOGY MATURITY

Behavior Modeling: Humans in/on the loop

- Support—than supplant—human (expert) analyses



Collaborative integration of human and machine intelligence

Behavioral Machine Intelligence: Human centered

COMPUTING

OF	human bio-behavioral data
FOR	meaningful analysis: timely decision making & intervention (action)
BY	collaborative integration of human expertise with automated processing: *support not supplant*

HUMANS

Shrikanth Narayanan and Panayiotis Georgiou. Behavioral Signal Processing: Deriving Human Behavioral Informatics from Speech and Language. Proceedings of IEEE. 101(5): 1203 - 1233, May 2013

Some Case Studies

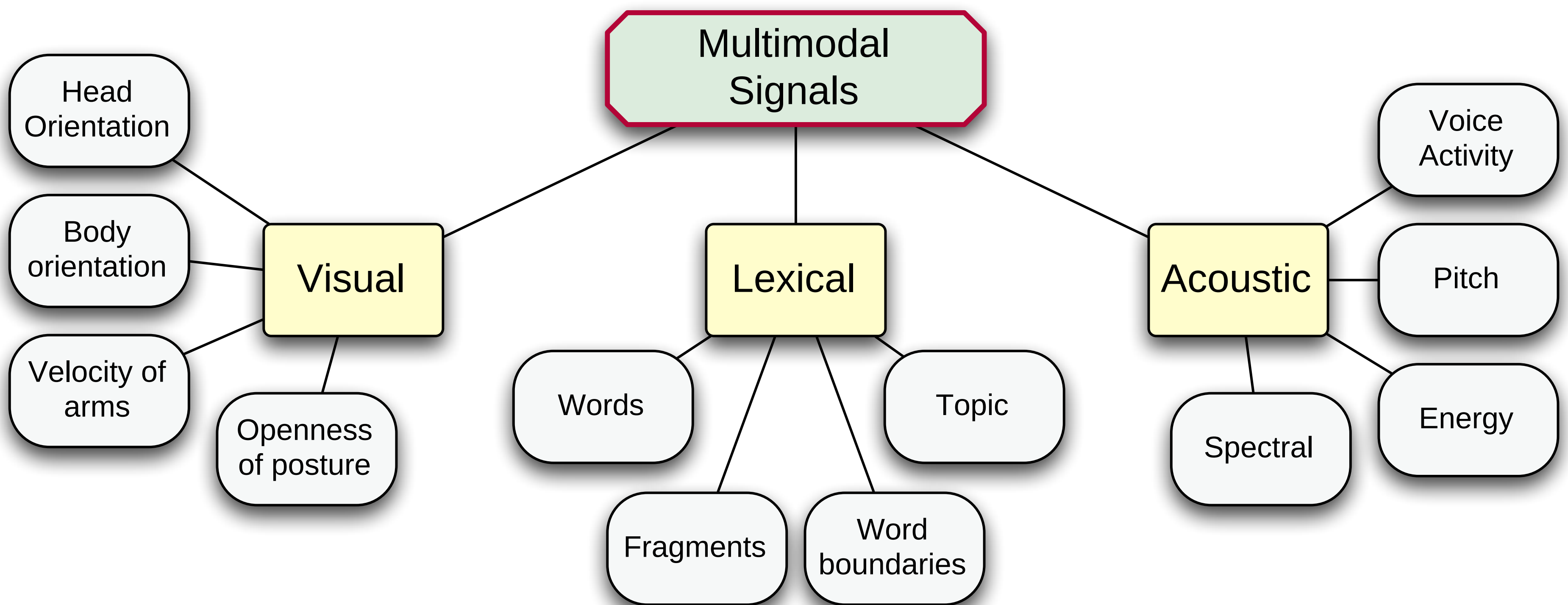
Modeling

Diagnostics

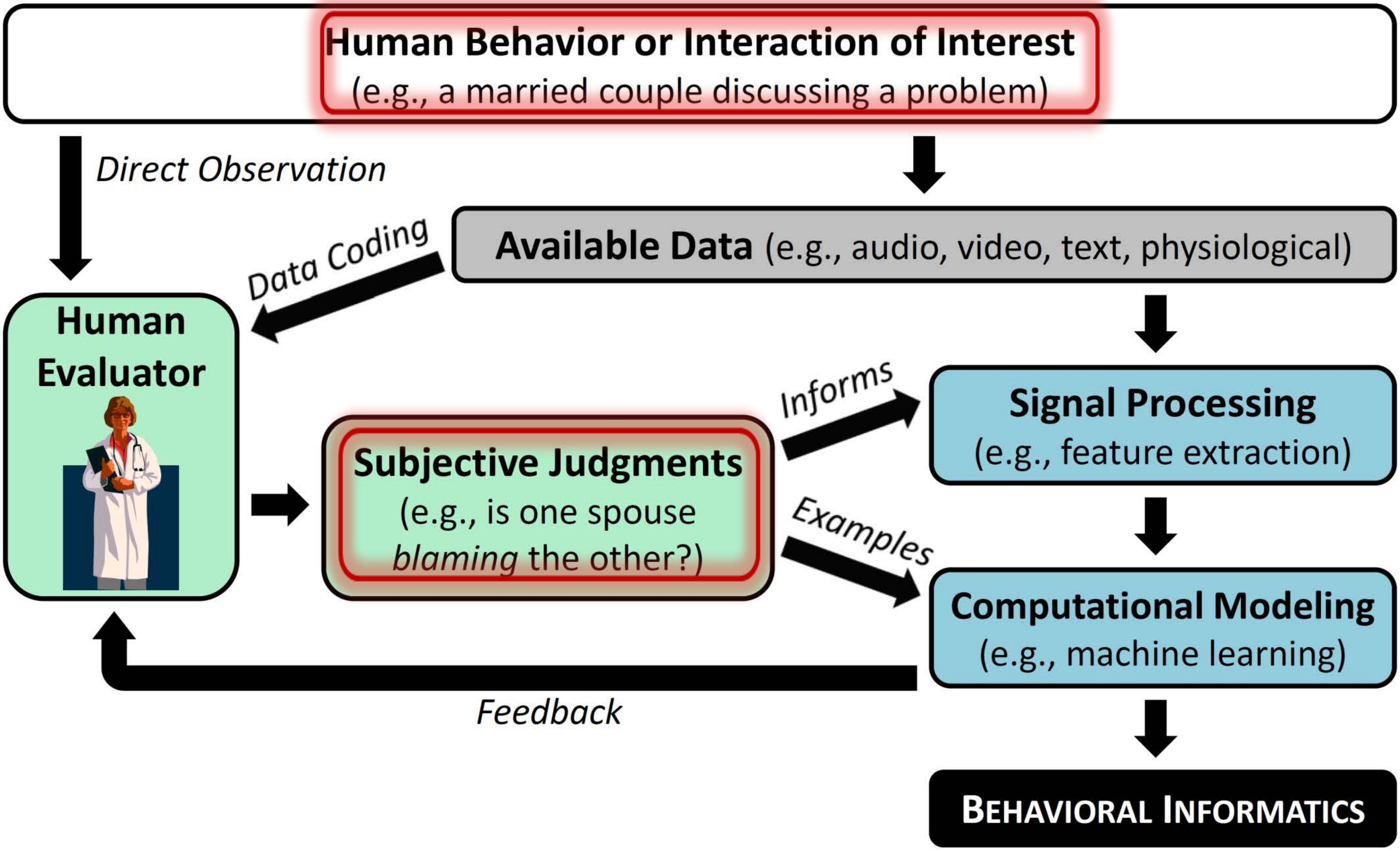
Intervention

Automatic Behavior Coding: Estimate behavioral codes from data

—*a multimodal machine intelligence exercise*



Couple Therapy Research



Dyadic Interactions of Distressed Couples

Characterizing affective dynamics, humor, blame patterns

"YOU WORK TOO MUCH…"

"SO HARD TO TALK ABOUT…"

"..IS REALLY HOUSEHOLD CHORES STUFF"

..TEMPER AND PATIENCE…"

CHRISTENSEN ET AL, JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY, 2004

Behavioral Coding by Human Experts

- ***Each* spouse evaluated by 3-4 trained coders**
  - 33 *session-level* codes (all on 1 to 9 scale)
  - No utterance- and turn-level ratings
  - Social Support Interaction Rating System
  - Couples Interaction Rating System
  - All evaluators underwent a training period to standardize the coding process

EXAMPLE CODING GOAL:
IS THE HUSBAND SHOWING ACCEPTANCE?" (SCALE 1-9)

FROM THE MANUAL:
"INDICATES UNDERSTANDING AND ACCEPTANCE OF PARTNER'S VIEWS, FEELINGS, AND BEHAVIORS. LISTENS TO PARTNER WITH AN OPEN MIND AND POSITIVE ATTITUDE. … "

- **Analyzed 6 codes for initial studies**
  - Level of **acceptance** ("*acc*")
  - Level of **blame** ("*bla*")
  - Global **positive affect** ("*pos*")
  - Global **negative affect** ("*neg*")
  - Level of **sadness** ("*sad*")
  - Use of **humor** ("*hum*")

Code	Code Correlation					Spouse Correlation	Agreement
	acc	*bla*	*pos*	*neg*	*sad*		
acc						0.647	0.751
bla	-0.80					0.470	0.788
pos	0.67	-0.54				0.667	0.740
neg	-0.77	0.72	-0.69			0.690	0.798
sad	-0.18	0.19	-0.18	0.36		0.315	0.722
hum	0.33	-0.20	0.47	-0.29	-0.15	0.787	0.755

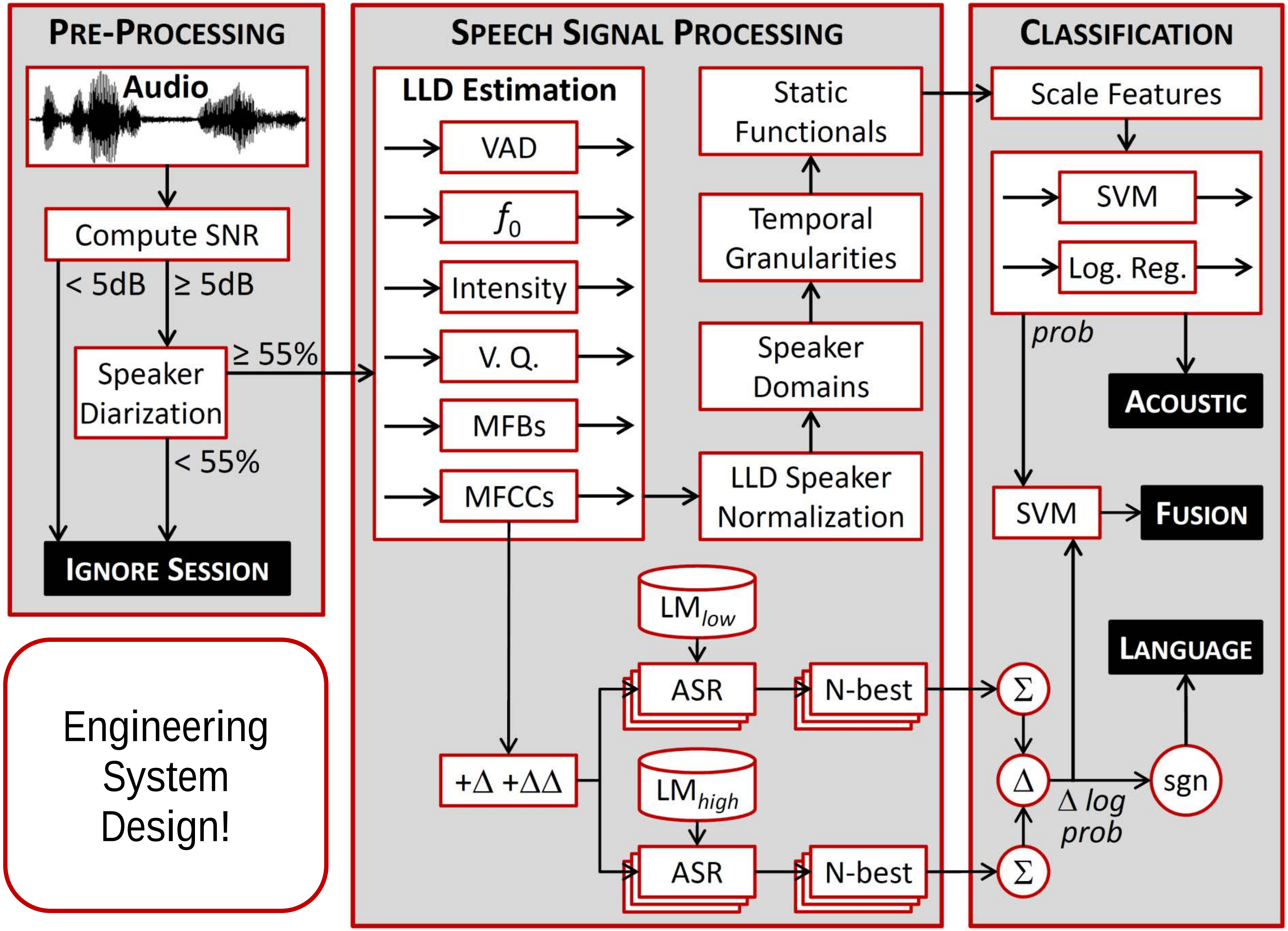
Domain Use case

Dyadic interaction of distressed couples

- **Real couples in 10-minute problem-solving interactions**
  - Longitudinal study at UCLA and UW [Christensen et al. 2004]
  - 134 distressed couples received couples therapy for 1 year
- **574 sessions (96 hours)**
  - Split-screen video (704x480 pixels, 30 fps)
  - Single channel of far-field audio
- **Data originally only intended for manual coding**
  - Recording conditions not ideal
  - Varied video angle, microphone placement, background noise

Methodology Pipeline

circa 2009

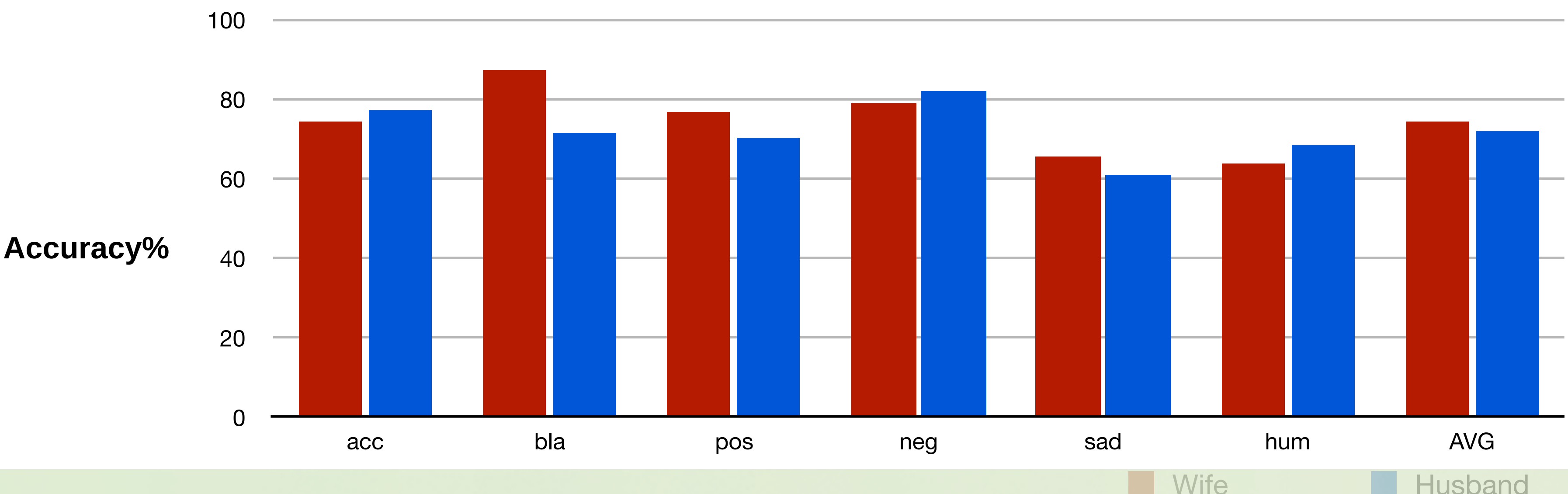


Engineering System Design!

Black, *et al.*, Classification of Blame in Married Couple's Interactions by Fusing Automatically Derived Speech and Language Cues, Interspeech, 2010 *circa 2009*

(Very) Simple Acoustic-feature based Behavior Estimation

- **Use of acoustic low-level descriptors (LLDs)**
  - Binary classification task
  - Linear-SVM
  - Global speaker-dependent cues capture evaluators' codes well
  - Capture relevant speech properties of spouses: every 10 ms:
    - Prosody (pitch, energy), spectral (MFCCs), voice quality (jitter, shimmer)
    - Separate features for each spouse (wife, husband)



M. BLACK, ET AL "AUTOMATIC CLASSIFICATION OF MARRIED COUPLES' BEHAVIOR USING AUDIO FEATURES" - INTERSPEECH 2010
M. BLACK, ET AL TOWARD AUTOMATING A HUMAN BEHAVIORAL CODING SYSTEM FOR MARRIED COUPLES' INTERACTIONS USING SPEECH ACOUSTIC FEATURES. SPEECH COMMUNICATION. 55(1):1-21, 2013

Lexical-information based Behavior Code Estimation

Partner	Transcript
H	WHAT DID I TELL YOU YOU CAN DO THAT AH AND EVERYTHING
W	BUT WHY DID YOU ASK THEN WHY DID TO ASK
H	AND DO IT MORE AND GET US INTO TROUBLE
W	YEAH WHY DID YOU ASK SEE MY QUESTION IS
H	MM HMMM
W	IF IF YOU TOLD ME THIS AND I AGREE I WOULD KEEP TRACK OF IT AND EVERYTHING
H	THAT'S THAT'S
W	**THAT'S AGGRAVATING VERY AGGRAVATING**
H	A BAD HABIT THAT
W	VERY AGGRAVATING
H	CAUSES YOU TO THINK THAT I DON'T TRUST YOU
W	THAT'S EXACTLY WHY THAT'S ABSOLUTELY THE WAY IT IS
H	AND IF I DON'T THE REASON FOR THAT IS AH
W	I DON'T CARE THE REASON YOU GET IT I GET IT TOO
H	THE REASON IS THE LONG TERM BAD PERFORMANCE
W	YEAH AND YOU KNOW WHY
H	MM HMMM
W	ALL YOU GET IS A NEGATIVE REACTION FROM ME

GEORGIOU, BLACK, LAMMERT, BAUCOM AND NARAYANAN. "THAT'S AGGRAVATING, VERY AGGRAVATING": IS IT POSSIBLE TO CLASSIFY BEHAVIORS IN COUPLE INTERACTIONS USING AUTOMATICALLY DERIVED LEXICAL FEATURES? PROCEEDINGS ACII, 2011

Informing experts

- Automated lexical analysis can inform experts
  - Example: Words that contributed to (correct) classification of a partner as "blaming"

Most blaming words in terms of discriminative contribution				Least blaming words in terms of discriminative contribution			
Word						me rob	Δ
YOU						.84	1.14
YOUR						.31	1.21
ME						.62	1.53
TELL						.32	1.55
ACCEPT						.07	1.56
CARING						.26	1.76
KITCHEN						.21	2.00
TOLD						.77	2.07
NOT	-40.32	-39.59	-0.73	WE	-29.39	-31.75	2.36
WHAT	-51.47	-50.77	-0.69	I	-99.92	-102.49	2.57
INTIMACY	-43.16	-42.53	-0.63	THAT	-91.30	-93.97	2.67
IT	-42.70	-42.18	-0.52	UM	-64.75	-70.76	6.01

High Blame		Low Blame	
word	*Δ log*	*word*	*Δ log prob*
YOU	-9.61	**UM**	6.01
YOUR	-4.06	**THAT**	2.67
ME	-2.53	**I**	2.57
TELL	-1.51	**WE**	2.36
ACCEPT	-1.45	**THINK**	2.07

Example Fusion Results:
Estimating “Blame”

Exploit complementary information from language and speech

Score-level fusion of classifiers using confidence scores

Classifier Type	Accuracy
Baseline Chance	50%
Language	75.4%
Acoustic	79.6%
Fusion	82.1%

- **REMARKS**

Lower performance of language classifier due to (our) ASR issues

Fusion advantageously uses language and acoustic information

Feasible to model high-level behaviors with automatically derived speech and language information

Many technical challenges & some approaches…

- Any single feature stream offers partial, noisy code information

➡ *Multimodal approach, Context sensitive learning*

- Not all portions of the feature stream are equally relevant in explaining an overall behavior description

➡ *Salient instances: Multiple instance learning*

- Behavior ratings are relative, often on an ordered scale

➡ *Ordinal regression*

- Behavior is a part of an interaction: mutual interlocutor dependency

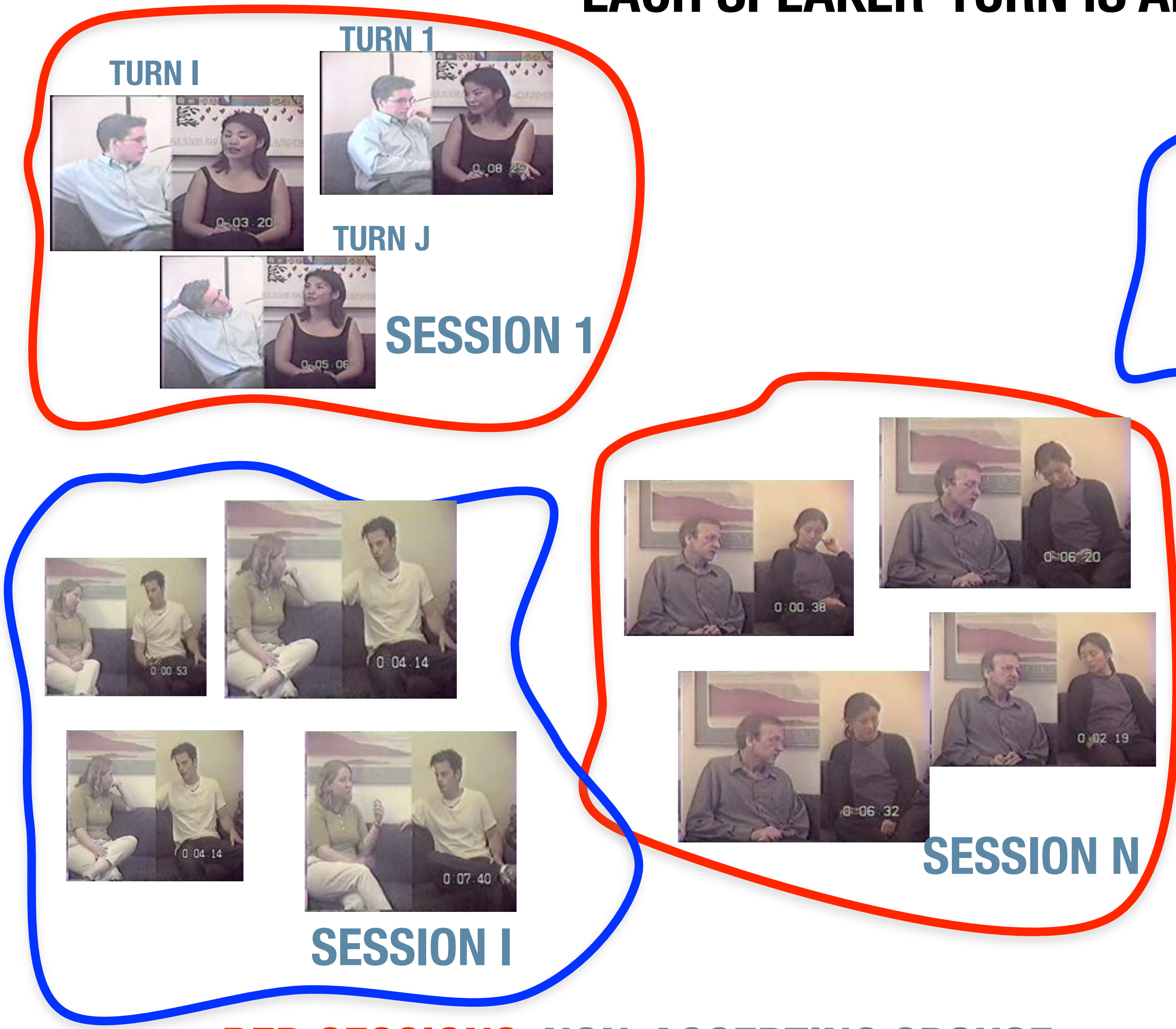
➡ *Models of entrainment*

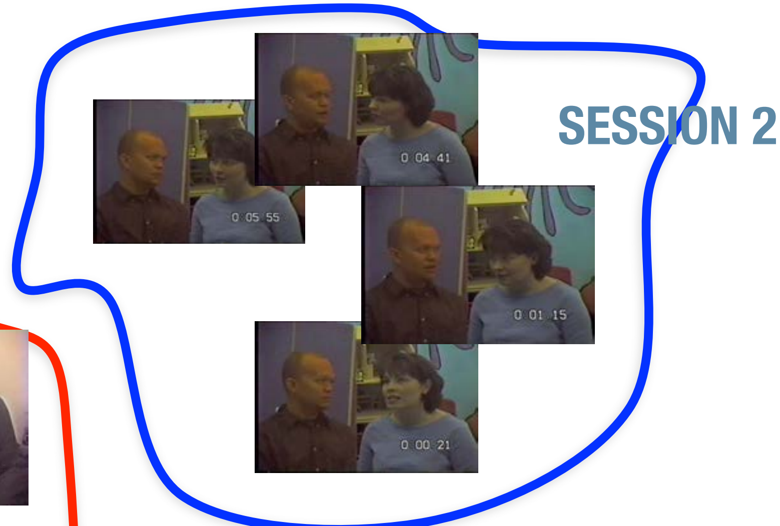
- Not all human observers/evaluators are equally reliable, and reliability is data dependent

➡ *Realistic models of human observers/evaluators*

Multiple Instance Learning

EACH SPEAKER-TURN IS AN INSTANCE (OF BEHAVIOR)





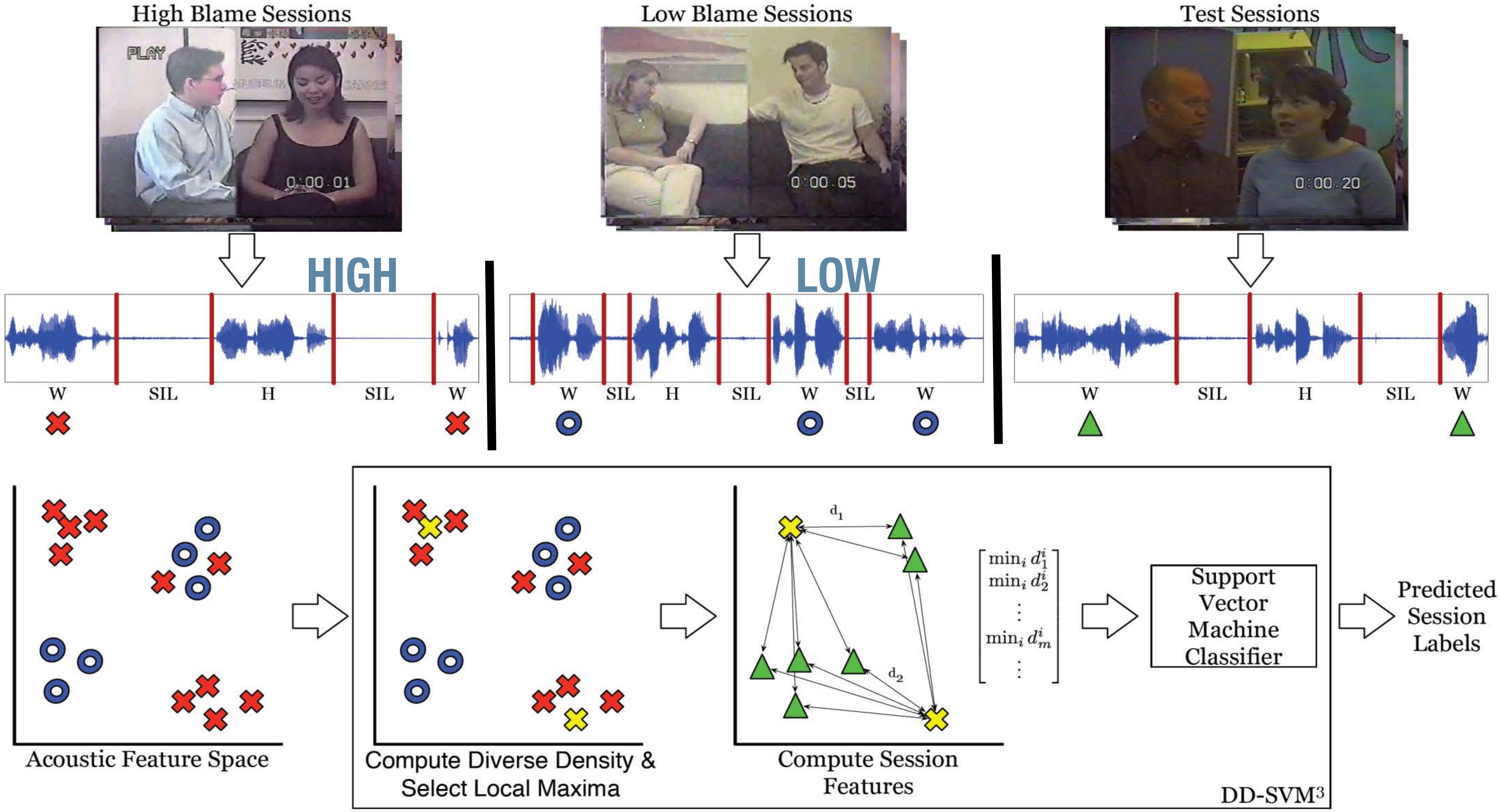
THE PROBLEM:
CAN WE IDENTIFY THE SPEAKER TURNS (INSTANCES) THAT ARE SALIENT, GIVEN THAT WE ONLY HAVE THE SESSION-LEVEL CODES?

RED SESSIONS: NON-ACCEPTING SPOUSE
BLUE SESSIONS: ACCEPTING SPOUSE

KATSAMANIS, GIBSON, BLACK, NARAYANAN, MULTIPLE INSTANCE LEARNING FOR CLASSIFICATION OF HUMAN BEHAVIOR OBSERVATIONS, ACII 2011
JAMES GIBSON, ATHANASIOS KATSAMANIS, FRANCISCO ROMERO, BO XIAO, PANAYIOTIS GEORGIOU, SHRIKANTH NARAYANAN. MULTIPLE INSTANCE LEARNING FOR BEHAVIORAL CODING. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, 2016

Saliency Detection with Multiple Instance Learning and Diverse Density SVM

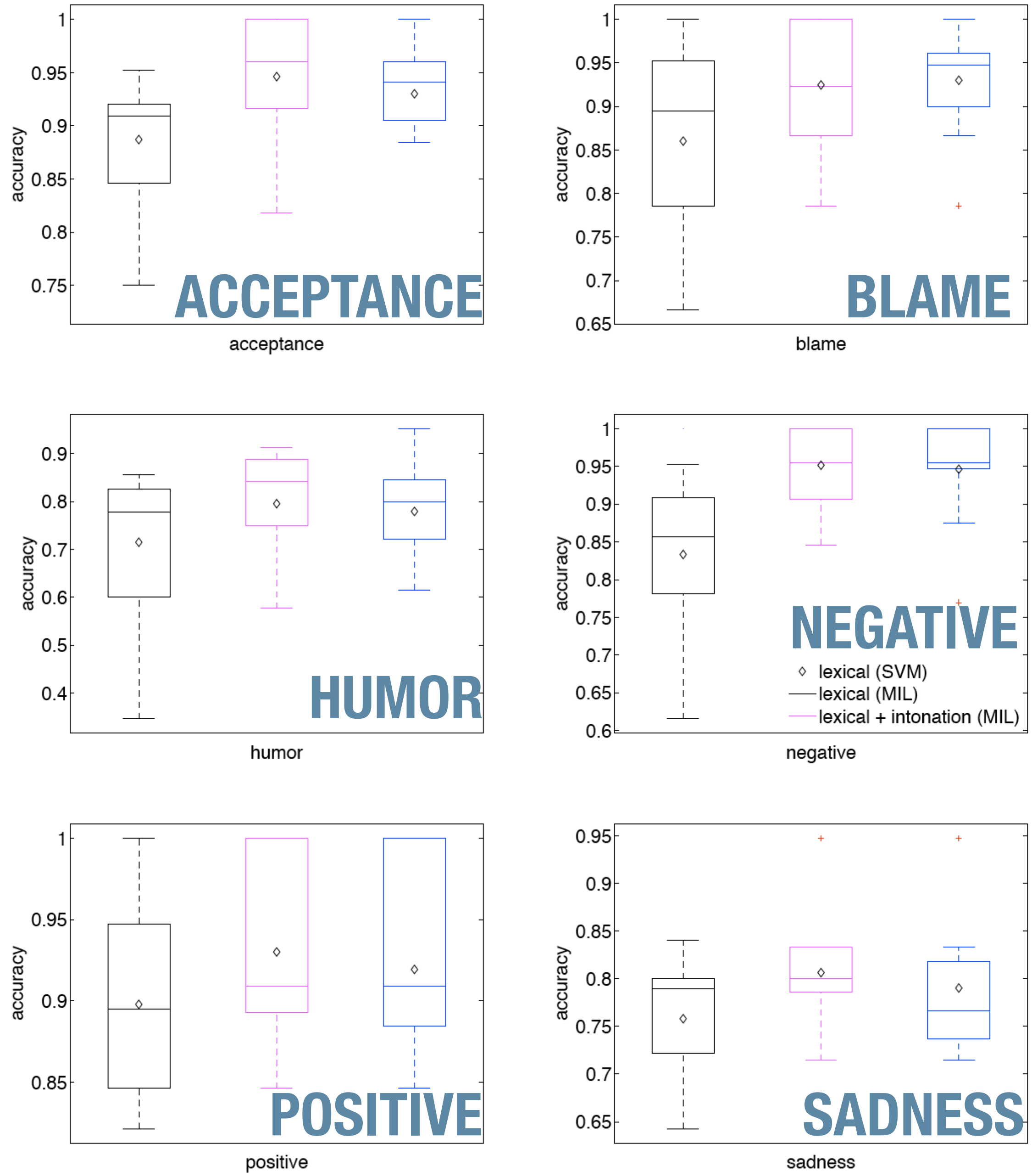
SALIENT PROTOTYPES: INSTANCES CLOSE TO POSITIVE BAGS AND FAR AWAY FROM NEGATIVE BAGS



A. KATSAMANIS, J. GIBSON, M. P. BLACK, AND S. S. NARAYANAN, "MULTIPLE INSTANCE LEARNING FOR CLASSIFICATION OF HUMAN BEHAVIOR OBSERVATIONS," IN: ACII, 2011.
J. GIBSON ET AL.,. MULTIPLE INSTANCE LEARNING FOR BEHAVIORAL CODING. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, 2016

Behavioral Coding Results: Instance Learning

Bag of Word Acoustic and Lexical features

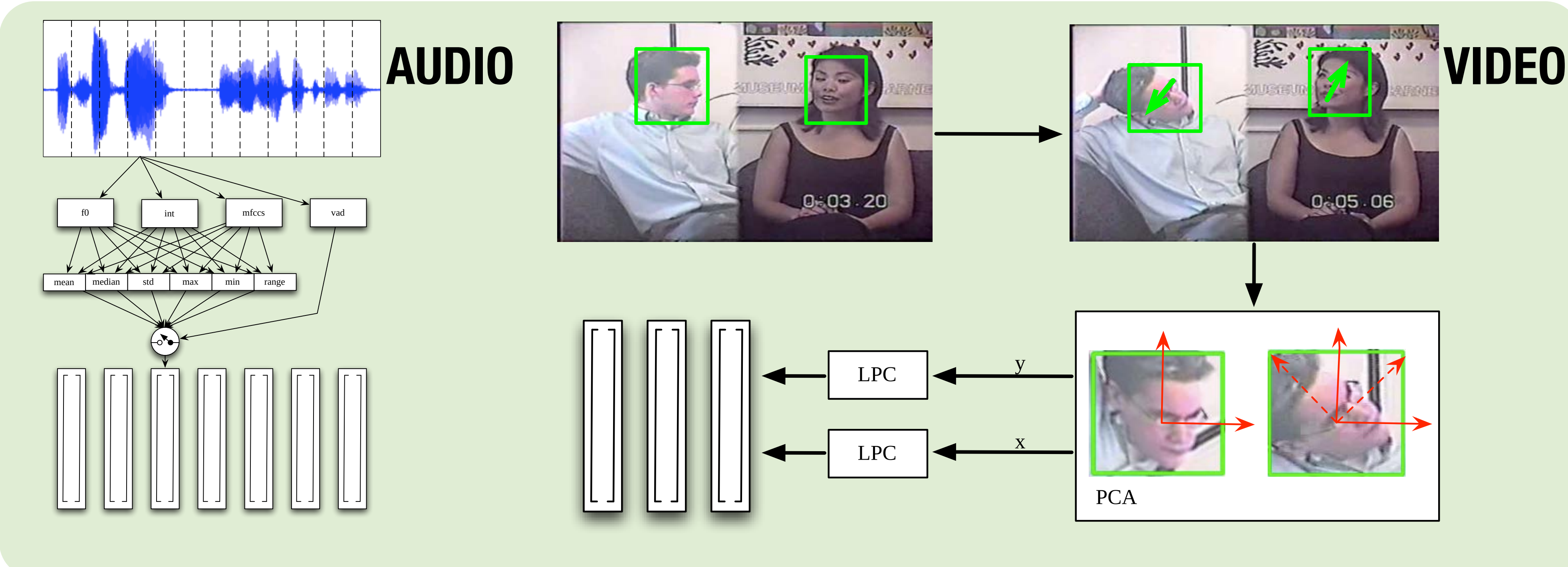


10-FOLD CROSS-VALIDATED RESULTS FOR SIX BEHAVIORAL CODES (HIGH VS LOW).

black boxes — baseline: Bag-of-words representation of the whole session (without exploiting saliency estimates)
red boxes — lexical + intonation (MIL)
blue boxes — lexical + intonation (MI)

SIGNIFICANT PERFORMANCE IMPROVEMENT WITH MULTIPLE INSTANCE LEARNING

Audio & Visual Salient Features



Classification accuracy (%) using audio, visual, and audio-visual fusion

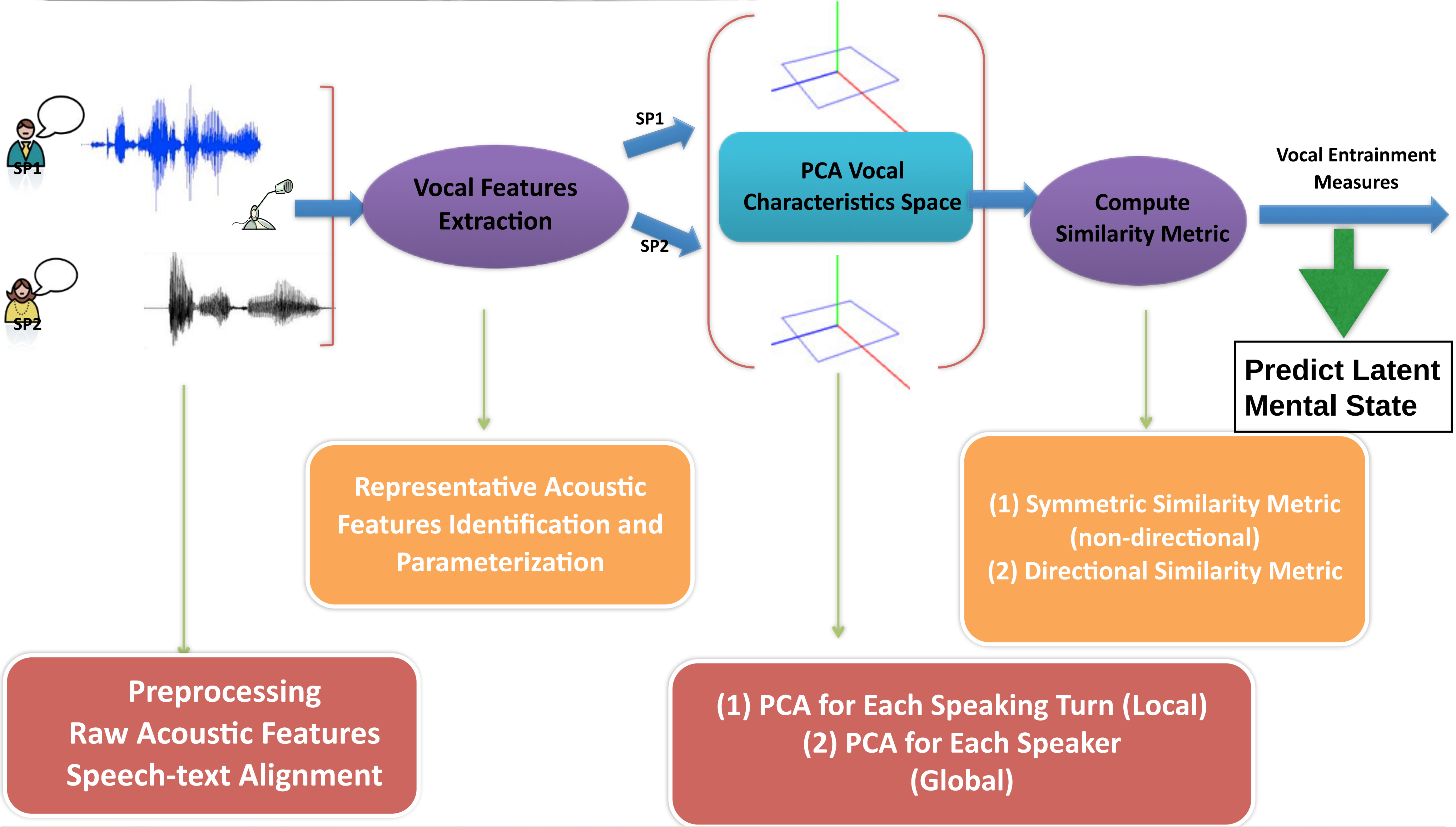
behavior	audio	visual	fusion	
			early	late
acceptance	70.5	62.5	64.3	72.3
blame	69.4	57.4	70.4	71.3

Late fusion improves accuracy
for classification of both behaviors

JAMES GIBSON, BO XIAO, PANAYIOTIS GEORGIOU, SHRIKANTH NARAYANAN, AN AUDIO-VISUAL APPROACH TO LEARNING SALIENT BEHAVIORS IN COUPLES' PROBLEM SOLVING DISCUSSIONS, IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA & EXPO (ICME), 2013

Computing Vocal Entrainment: A novel measure

“HOW MUCH DO TWO PEOPLE *SYNCHRONIZE* IN A CONVERSATION?”

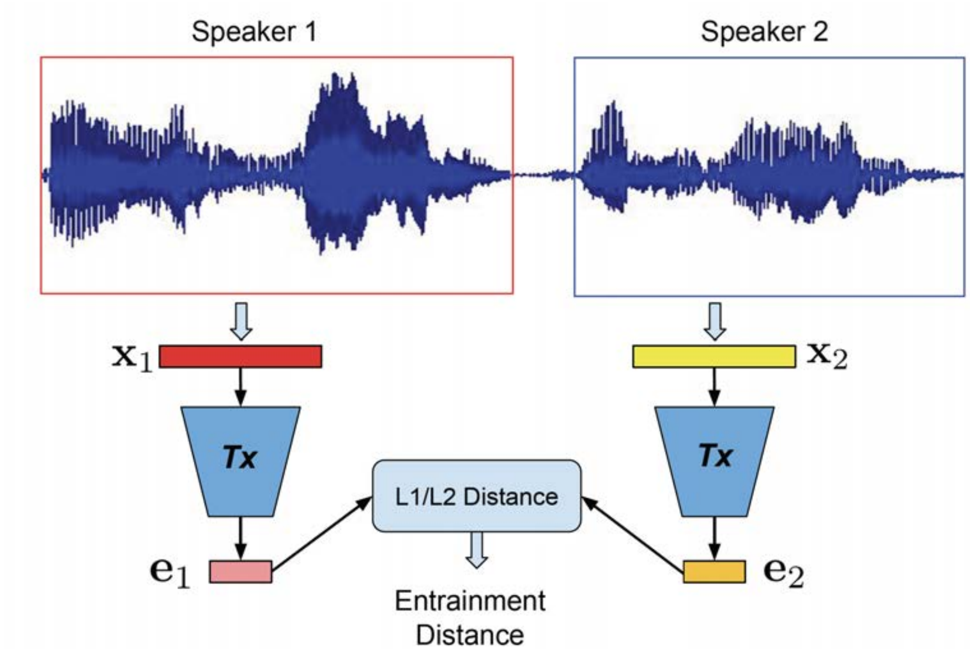


CHI-CHUN LEE, ET AL. COMPUTING VOCAL ENTRAINMENT: A SIGNAL-DERIVED PCA-BASED QUANTIFICATION SCHEME WITH APPLICATION TO AFFECT ANALYSIS IN MARRIED COUPLE INTERACTIONS. COMPUTER, SPEECH, AND LANGUAGE. 28(2): 518-539, MARCH 2014

MD NASIR, BRIAN BAUCOM, SHRIKANTH NARAYANAN, PANAYIOTIS GEORGIOU. MODELING VOCAL ENTRAINMENT IN CONVERSATIONAL SPEECH USING DEEP UNSUPERVISED LEARNING. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 2020

Deep Learning for Modeling Vocal Entrainment

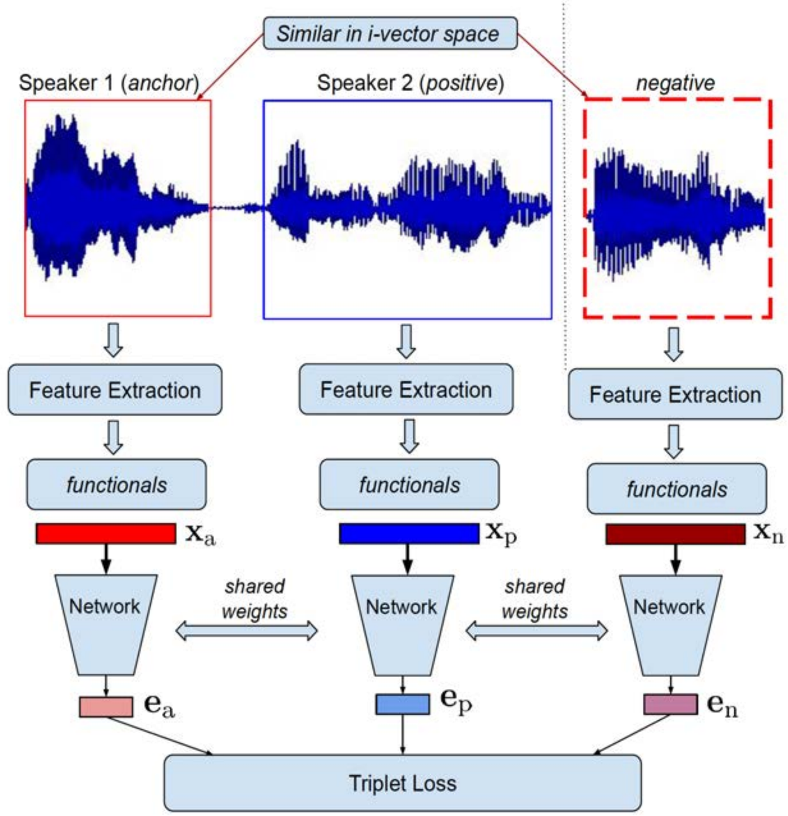
- Transform acoustic features of speaker turns to embeddings
- Obtain a minimal representation of:
  - information that can be 'transferred' across interlocutors
  - related to entrainment
- Compute entrainment distance in the embedding space
- A number of distances proposed under this framework using different neural network modeling approaches:
  - NED, TNED, iTNED
- iTNED seems to perform the best in the experiments in association with
  - couples therapy codes (agreement and blame)
  - couples therapy outcome
  - Emotional bond in Suicide risk assessment interviews



MD NASIR, BRIAN BAUCOM, SHRIKANTH NARAYANAN, PANAYIOTIS GEORGIOU. MODELING VOCAL ENTRAINMENT IN CONVERSATIONAL SPEECH USING DEEP UNSUPERVISED LEARNING. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 2020

i-vector-based Triplet Network Entrainment Distance (iTNED)

- ***Triplets***
  - **Anchor**: previous speaker turn
  - **Positive:** next speaker turn
  - **Negative**: strategically chosen non-consecutive turn using proposed i-vector based sampling strategy
- Minimizing triplet loss
  - minimizes anchor-positive distance, thus preserves entrainment information
  - maximizes anchor-negative distance, thus reduces nuisance factors (speaker and channel characteristics)
- **iTNED:** Distance measured in the embedding (last layer of the trained network)



Computing Multi/Cross-modal Entrainment & Synergy

- Computational models of synchrony between head, hand and body gestures and vocal patterns
- Use to
  - characterize behavioral constructs e.g., ***approach-avoidance, affect, empathy,..***
  - predict the behavior of the other interactant

ANGELIKI METALLINOU, ATHANASIOS KATSAMANIS AND SHRIKANTH NARAYANAN. TRACKING CONTINUOUS EMOTIONAL TRENDS OF PARTICIPANTS DURING AFFECTIVE DYADIC INTERACTIONS USING BODY LANGUAGE AND SPEECH INFORMATION. JOURNAL IMAGE AND VISION COMPUTING. 31(2): 137-152, FEBRUARY 2013

ZHAOJUN YANG AND SHRIKANTH NARAYANAN. MODELING DYNAMICS OF EXPRESSIVE BODY GESTURES IN DYADIC INTERACTIONS. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 8(3): 369 - 381, JULY 2017

ANGELIKI METALLINOU, ZHAOJUN YANG, CHI-CHUN LEE, CARLOS BUSSO, SHARON CARNICKE AND SHRIKANTH NARAYANAN. THE USC CREATIVEIT DATABASE OF MULTIMODAL DYADIC INTERACTIONS: FROM SPEECH AND FULL BODY MOTION CAPTURE TO CONTINUOUS EMOTIONAL ANNOTATIONS. JOURNAL OF LANGUAGE RESOURCES AND EVALUATION. PP. 1-25, 2015

ZHAOJUN YANG, ANGELIKI METALLINOU AND SHRIKANTH S. NARAYANAN. ANALYSIS AND PREDICTIVE MODELING OF BODY LANGUAGE BEHAVIOR IN DYADIC INTERACTIONS FROM MULTIMODAL INTERLOCUTOR CUES. IEEE TRANSACTIONS ON MULTIMEDIA. 16(6): 1766-1778, OCTOBER 2014.

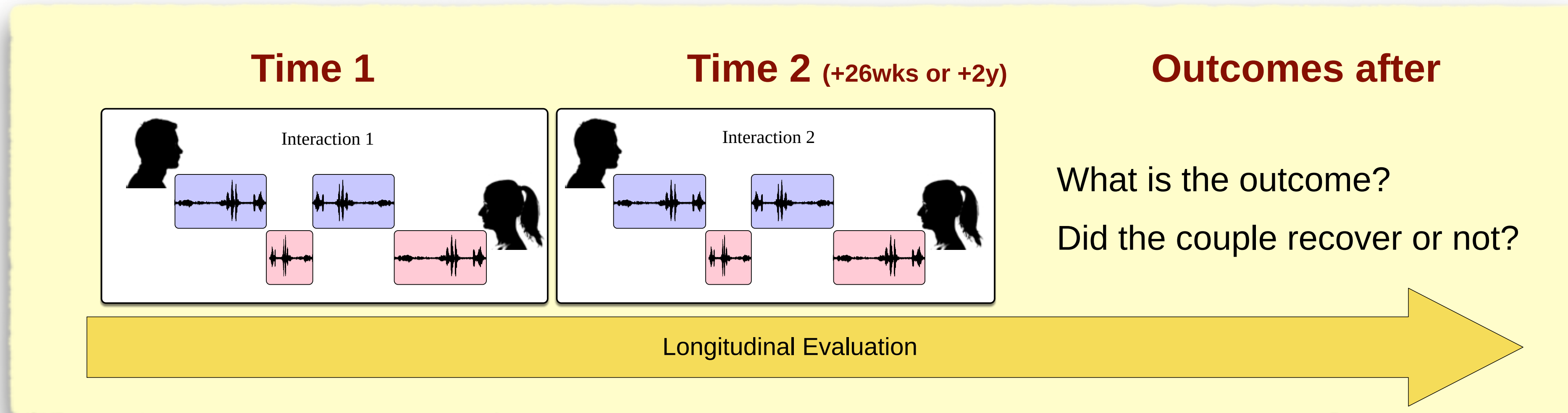
BO XIAO, PANAYIOTIS GEORGIOU, BRIAN BAUCOM, SHRIKANTH NARAYANAN. HEAD MOTION SYNCHRONY AND ITS CORRELATION TO AFFECTIVITY IN DYADIC INTERACTIONS. IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA & EXPO, 2013

BO XIAO, PANAYIOTIS GEORGIOU, BRIAN BAUCOM AND SHRIKANTH S. NARAYANAN. HEAD MOTION MODELING FOR HUMAN BEHAVIOR ANALYSIS IN DYADIC INTERACTION. IEEE TRANSACTIONS ON MULTIMEDIA. 17(7): 1107-1119, JULY 2015

Relationship Outcomes: Effect of therapy?

Md Nasir, Brian Baucom, Panayiotis Georgiou, Shrikanth S. Narayanan. Predicting Couple Therapy Outcomes Based on Speech Acoustic Features. *PLoS ONE.* 12(9): e0185123, 2017.

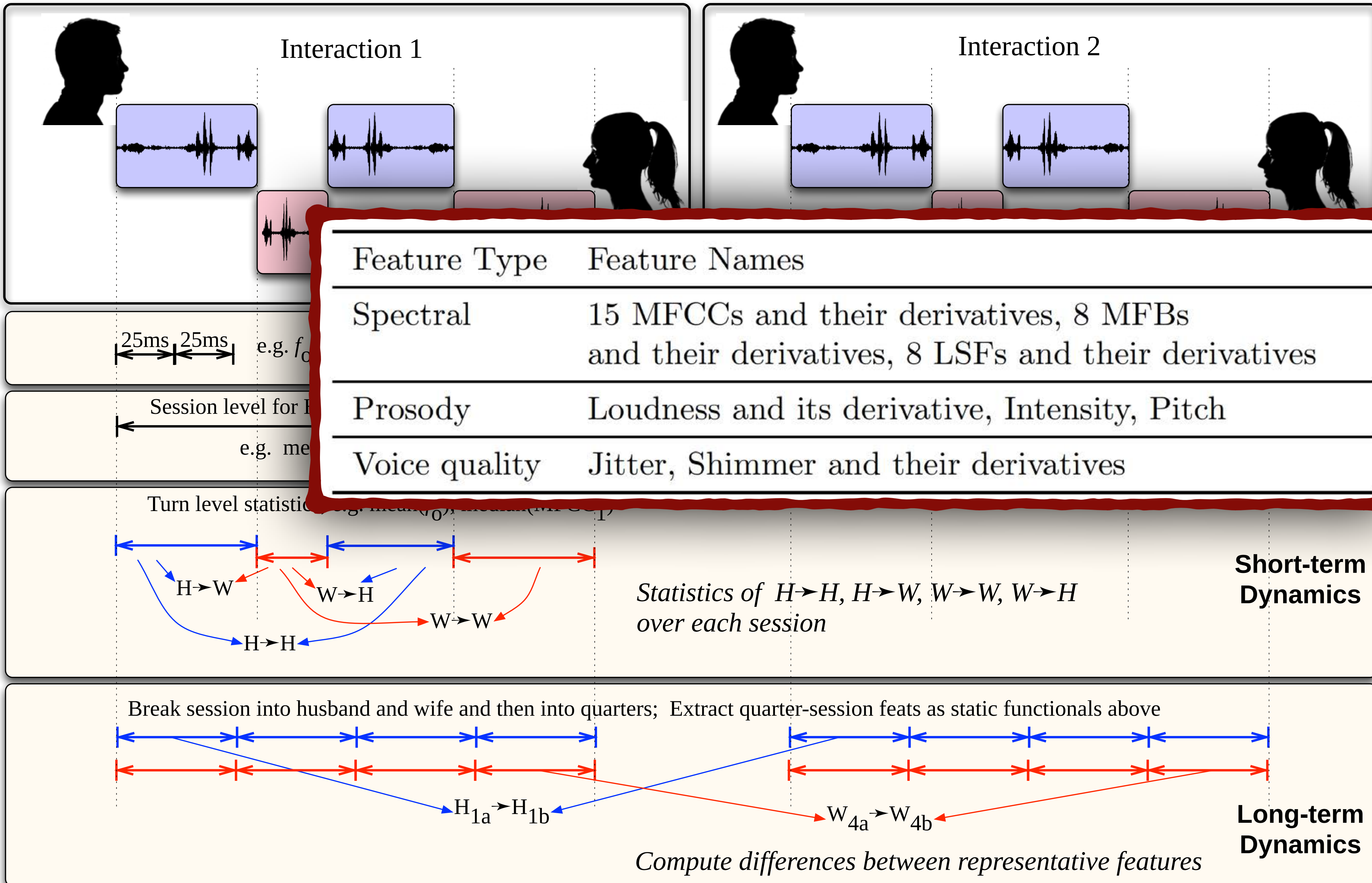
- Outcome prediction
  - Experimental setup/Task
  - Features
  - Results



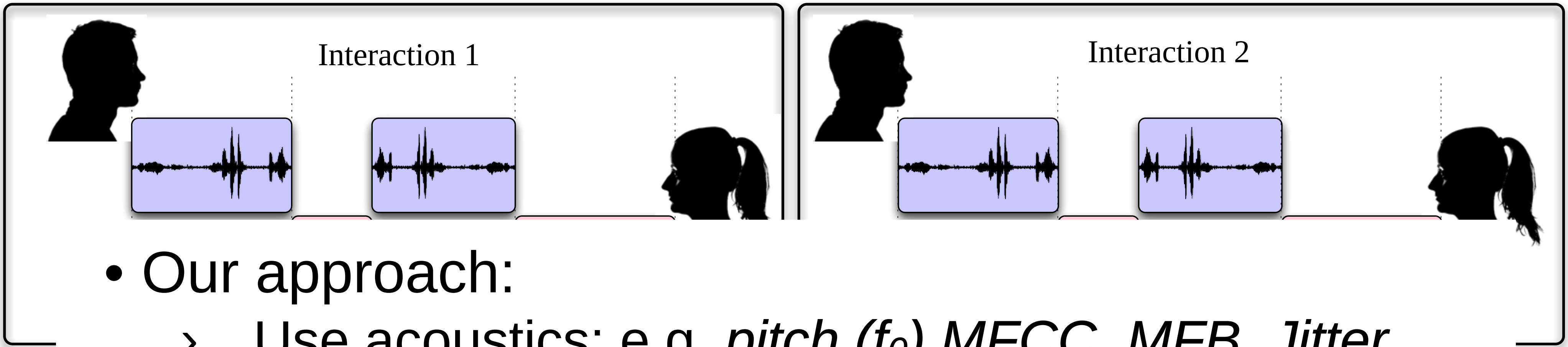
- › ratings 1&2&3 *vs* 4
- Experiment 2:
  - › No (or incomplete) recovery into finer levels,
  - › i.e., rating **1** *vs* rating **2** *vs* rating **3**
- Experiment 3:
  - › All possible outcomes

Outcome	Decline	No Change	Partial Recovery	Recovery
Rating	1	2	3	4
Count	12	26	34	67

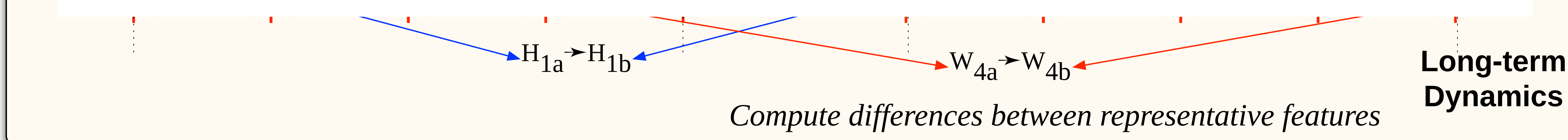
- Outcome prediction
    - Experimental setup/Task
    - Features
    - Results

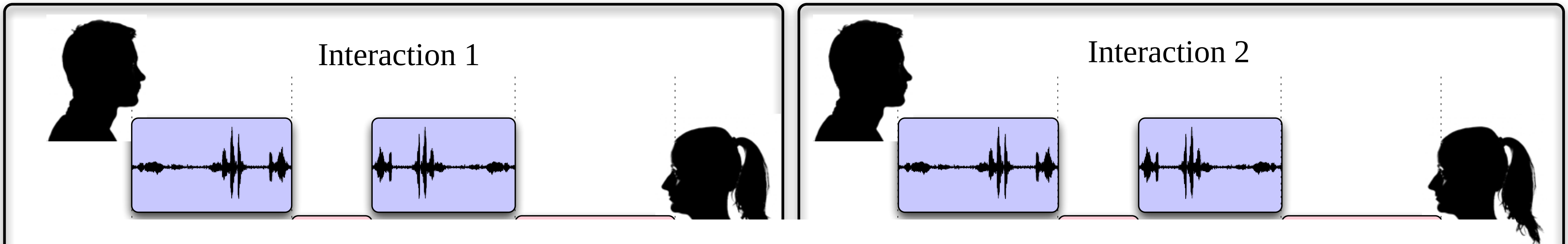


Feature Type	Feature Names
Spectral	15 MFCCs and their derivatives, 8 MFBs and their derivatives, 8 LSFs and their derivatives
Prosody	Loudness and its derivative, Intensity, Pitch
Voice quality	Jitter, Shimmer and their derivatives

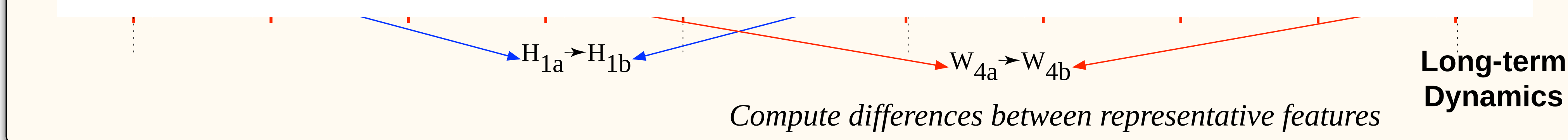


- Our approach:
  - › Use acoustics: e.g. *pitch (f_0) MFCC, MFB, Jitter, Loudness*
  - › Extract those for husband, wife separately
  - › both sessions; 4 interactions

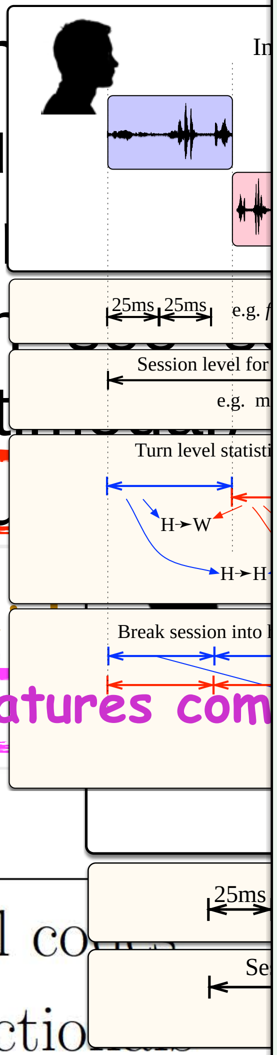




- Describe ***Dynamics:*** how each person's acoustics in every turn
  - › relate to partner
  - › relate to self



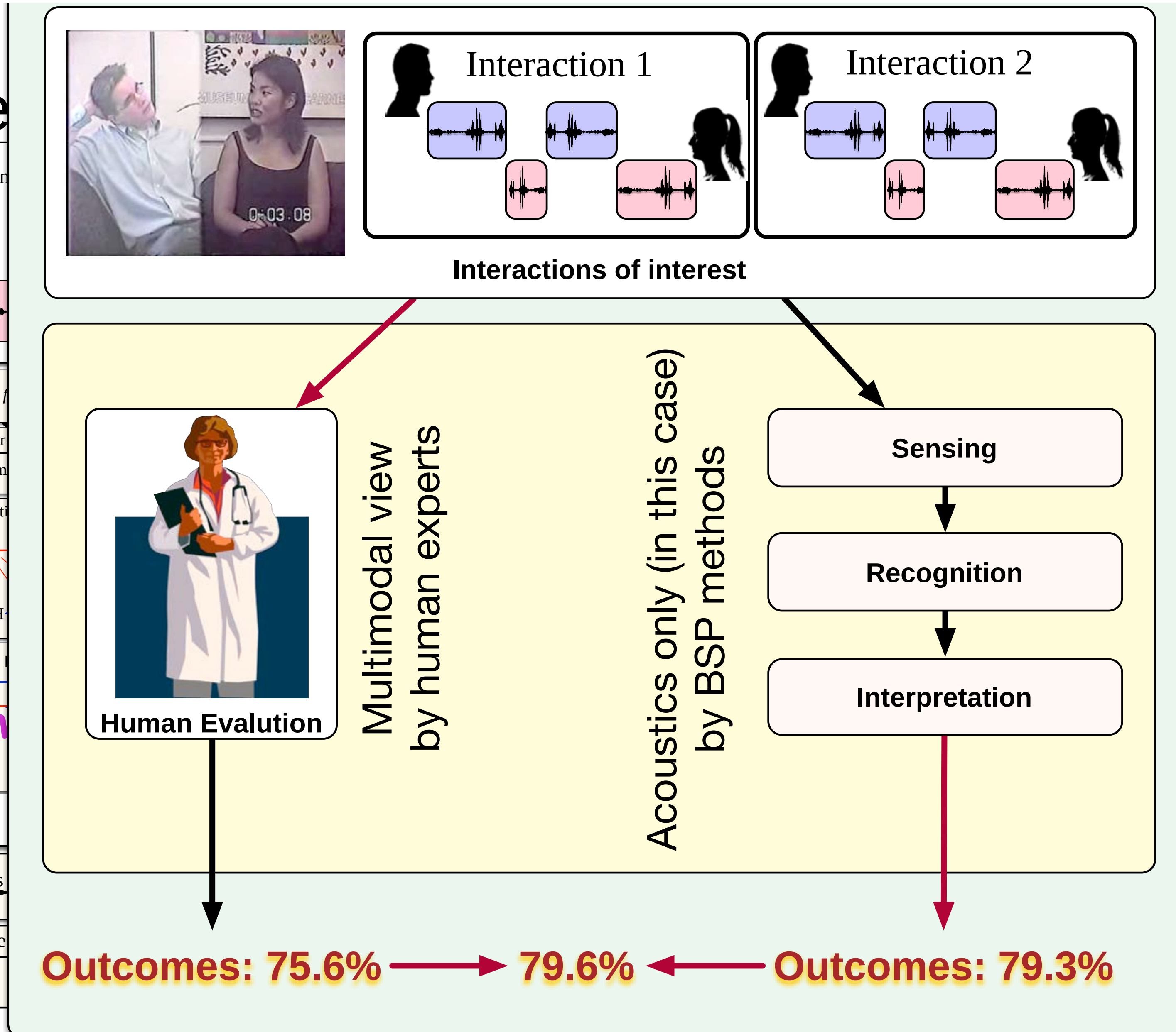
- Human code-de
worse th
- Static fu
behavio
  - › Both
mult
- › Introdu



Static

Dynamic: Features com

Featureset							
Chance							
Behavioral co							
Static functio							
Dynamic functionals	6696	78.9	7.6	71.1	12.8	61.5	12.3
Acoustic (all functionals)	10248	79.3	10.2	72.6	13.0	**64.1**	12.8
All features	9144	**79.6**	7.4	**74.6**	12.6	64.1	13.2



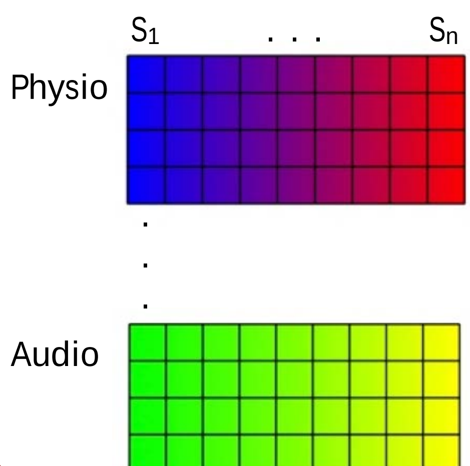
Beyond traditional clinical settings? bringing the care directly to the individual

- Home, work place,…
  - remote health tracking
  - stress regulation at work
  - conflict at home

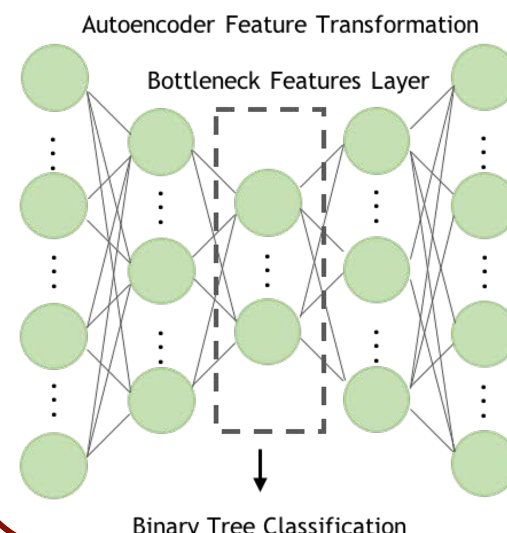
Conflict detection system

Data collection methods

Multimodal feature design



Multimodal fusion



Clinical state tracking in serious mental illness through computational analysis of speech

A. AREVIAN, D. BONE, N. MALANDRAKIS, V. MARTINEZ, K. WELLS, D. MIKLOWITZ, S. NARAYANAN. CLINICAL STATE TRACKING IN SERIOUS MENTAL ILLNESS THROUGH COMPUTATIONAL ANALYSIS OF SPEECH. PLOS ONE. 15(1): E0225695, 2020

Can machine learning help us scale better and more personalized mental health services?

In 2019, nearly one in five U.S. adults live with a mental illness (NIMH, 2021)

- *Only 23 million (44.8%) received any type of mental health services*
- Serious mental illness affect an estimated 13.1 million adults (5.2%)

Speech and language may provide a way into an individual's underlying neuro-psychiatric states

Photo by Toa Heftiba on Unsplash

Longitudinal study of 4–14 months of speech samples from adults with serious mental illness

Participants (n=47) with serious mental illness answered 3 open ended prompts (2-3 mins call)

- ***1101 phone calls, 117 hours of speech***
- ***Expert providers (N=13) assessed mental health state (1→10)***

Speech Features

- ***Acoustic: pitch, intonation, inter-word pause***
- ***Lexical: emotion, complexity, affect, concreteness,..***

Lexical and acoustic features can be used to track changes in mental health states over time

Feature	Set	Functional	Correlation
Negative emotion words	LIWC	% words	-0.36
Positive emotion words	LIWC	% words	+0.34
Valence	Lexical Norms	Mean	+0.32
Negative	Lexical Norms	Mean	-0.32
Positive	Lexical Norms	Max	+0.26
Difficulty of words	Complexity	Mean	+0.21
Religious words	LIWC	% words	+0.20
Gender ladenness	Lexical Norms	Min	-0.20
Arousal	Lexical Norms	Min	-0.19
2^{nd} vocal formant	Acoustics	Mean	+0.18
Sad words	LIWC	% words	-0.16
Coherence (latent semantic analysis)	Complexity	Stdv.	+0.16
SMOG Index	Complexity	Mean	+0.14
Harmonicity	Acoustics	Median	-0.13
Assent	LIWC	% words	+0.12

LIWC, Linguistic Inquiry of Word Count toolkit; SMOG, Subjective Measure of Gobbligook Index.

https://doi.org/10.1371/journal.pone.0225695.t002

Provider ratings associated with ↓negative emotion, ↑positive emotion, ↑complex language, ↑readability, ↑semantic coherence, and ↓harmonicity

Personalized models achieved the highest correlation (ρ = 0.78)

THE USC COUPLE MOBILE SENSING PROJECT
http://homedata.github.io/

Multimodal ambulatory detection of relationship conflict

self-reports
- ✓ mood
- ✓ quality of interactions

context and interaction
- ✓ GPS
- ✓ activity count
- ✓ body temperature
- ✓ alcohol/caffeine/drugs

electrodermal activity
- ✓ skin conductance level
- ✓ skin conductance response

language use
- ✓ linguistic constructs
- ✓ psychological factors
- ✓ personal concern
- ✓ paralinguistic

electrocardiogram
- ✓ heart rate
- ✓ heart rate variability

physiological synchrony
- ✓ joint sparse representation
- ✓ multiple time scales

acoustic analysis
- ✓ pitch (F0)
- ✓ intensity

Unweighted classification accuracy up to 81% and 86% for females and males

Adela C. Timmons, Theodora Chaspari, Sohyun C. Han, Laura Perrone, Shrikanth S. Narayanan, and Gayla Margolin. Multimodal Detection of Conflict in Couples Using Wearable Technology. IEEE Computer. Special Issue on Quality-of-Life Technologies. March 2017.

USC Couple Mobile Sensing Project

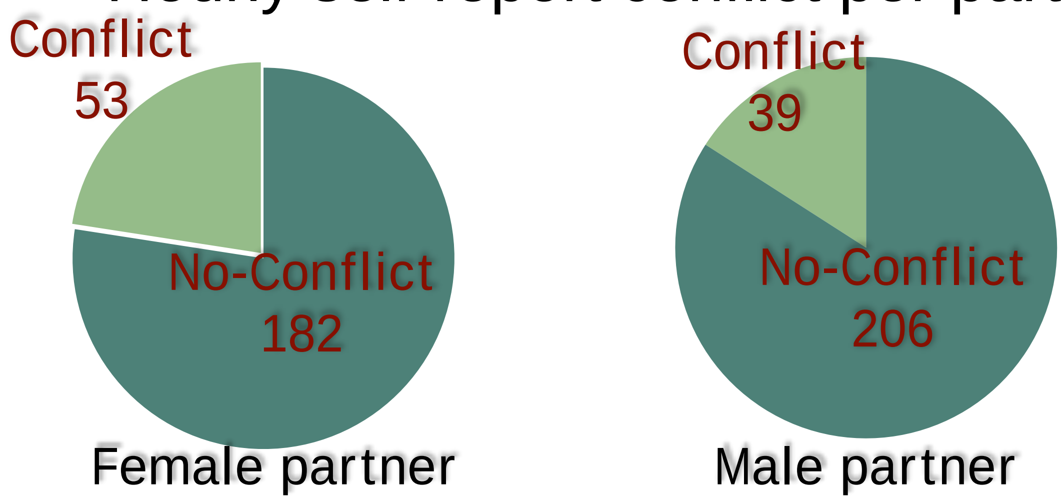
THE USC COUPLE MOBILE SENSING PROJECT

http://homedata.github.io/

Data source
- ✓ young-adult dating couples
- ✓ 34 couples
- ✓ 22.5 years average

Collection procedures
- ✓ 1 smartphone: self-reports, GPS, audio
- ✓ 2 wearable sensors: EDA, ECG

Conflict labels
- ✓ Hourly self-report conflict per partner



Relationship Conflict Prediction

THE USC COUPLE MOBILE SENSING PROJECT
http://homedata.github.io/

Unweighted classification accuracy (%)

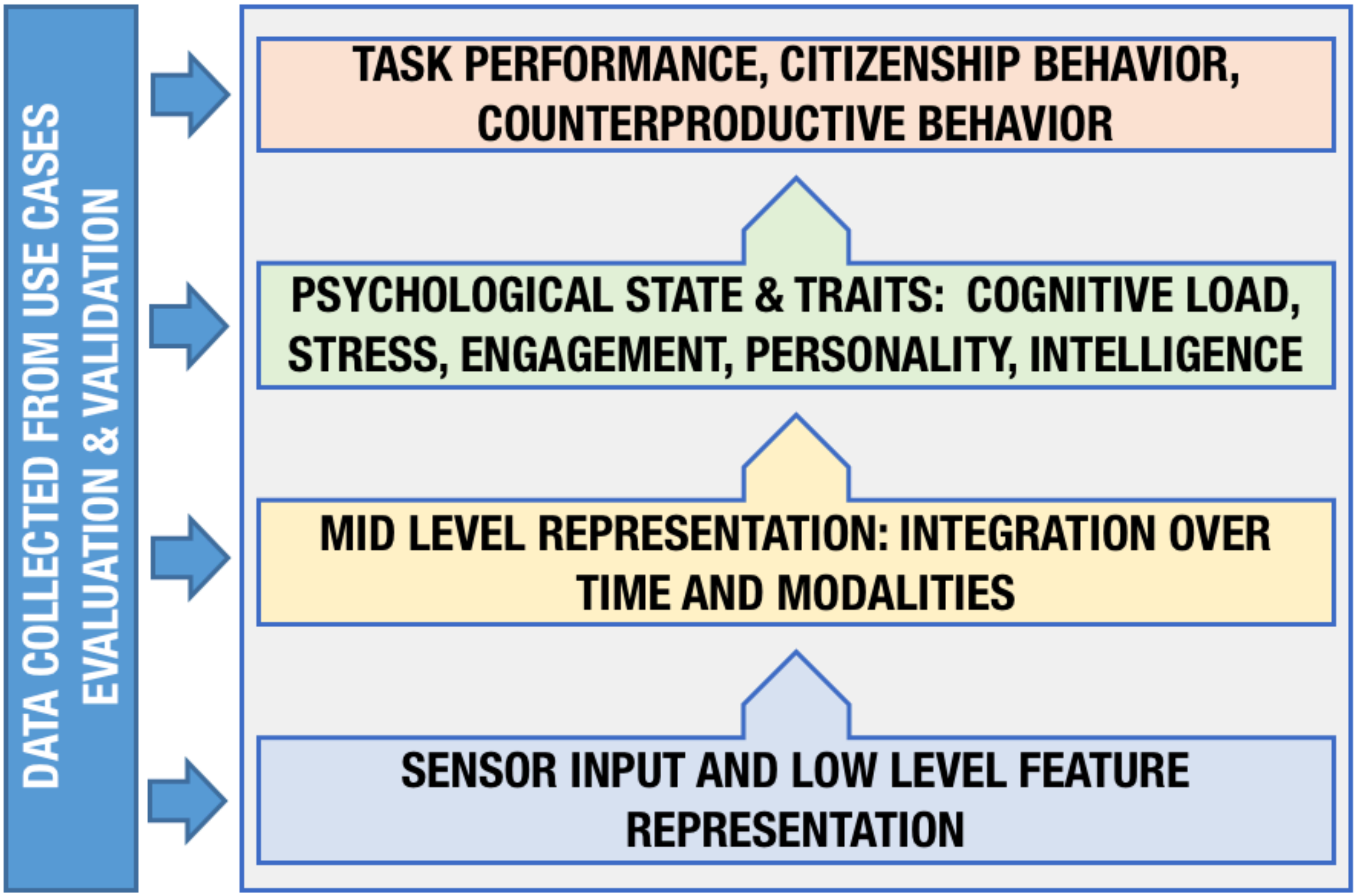
Features	System 1 Decision-level fusion		System 2 Feature-level fusion	
	Female	Male	Female	Male
Self-report	70.4**	61.5*	58.3	67.9**
Multimodal	73**	76.9**	74.2**	76.9**
Self-report+Multimodal	**78.3****	**81.2****	**79.4****	**86.3****

*p<0.05, **p<0.01 (UA significantly higher than 50% chance)

A. Timmons, T. Chaspari, S. C. Han, L. Perrone, S. Narayanan, and G. Margolin. Multimodal Detection of Conflict in Couples Using Wearable Technology. IEEE Computer. 50(3): 50-59, March 2017.
A. Timmons, B. Baucom, S. Han, L. Perrone, T. Chaspari, S. Narayanan, and G. Margolin. New Frontiers in Ambulatory Assessment: Big Data Methods for Capturing Couples' Emotions, Vocalizations, and Physiology in Daily Life. Social Psychological and Personality Science. 2017

TILES: Tracking individual performance with sensors (at work place)

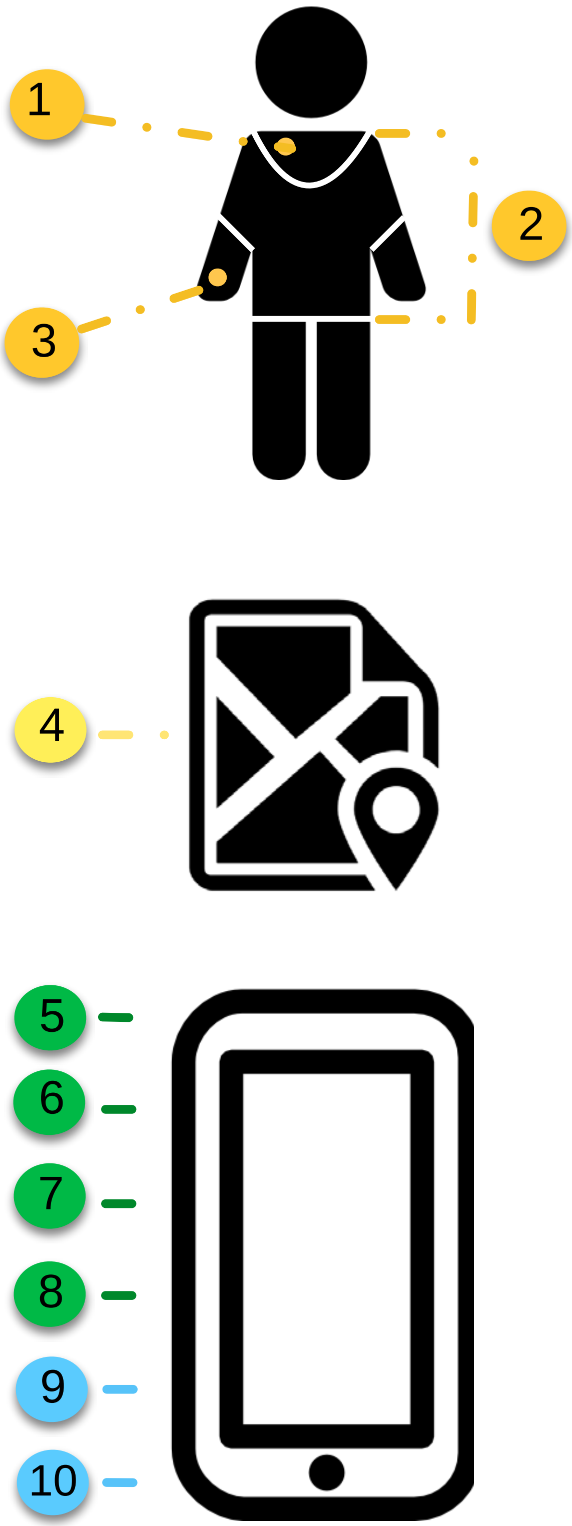
End-to-end bio-behavioral platform for individualized performance assessment from sensor data



"Be Well--Do Well"

IARPA MOSAIC Program

TILES Data Collection Streams

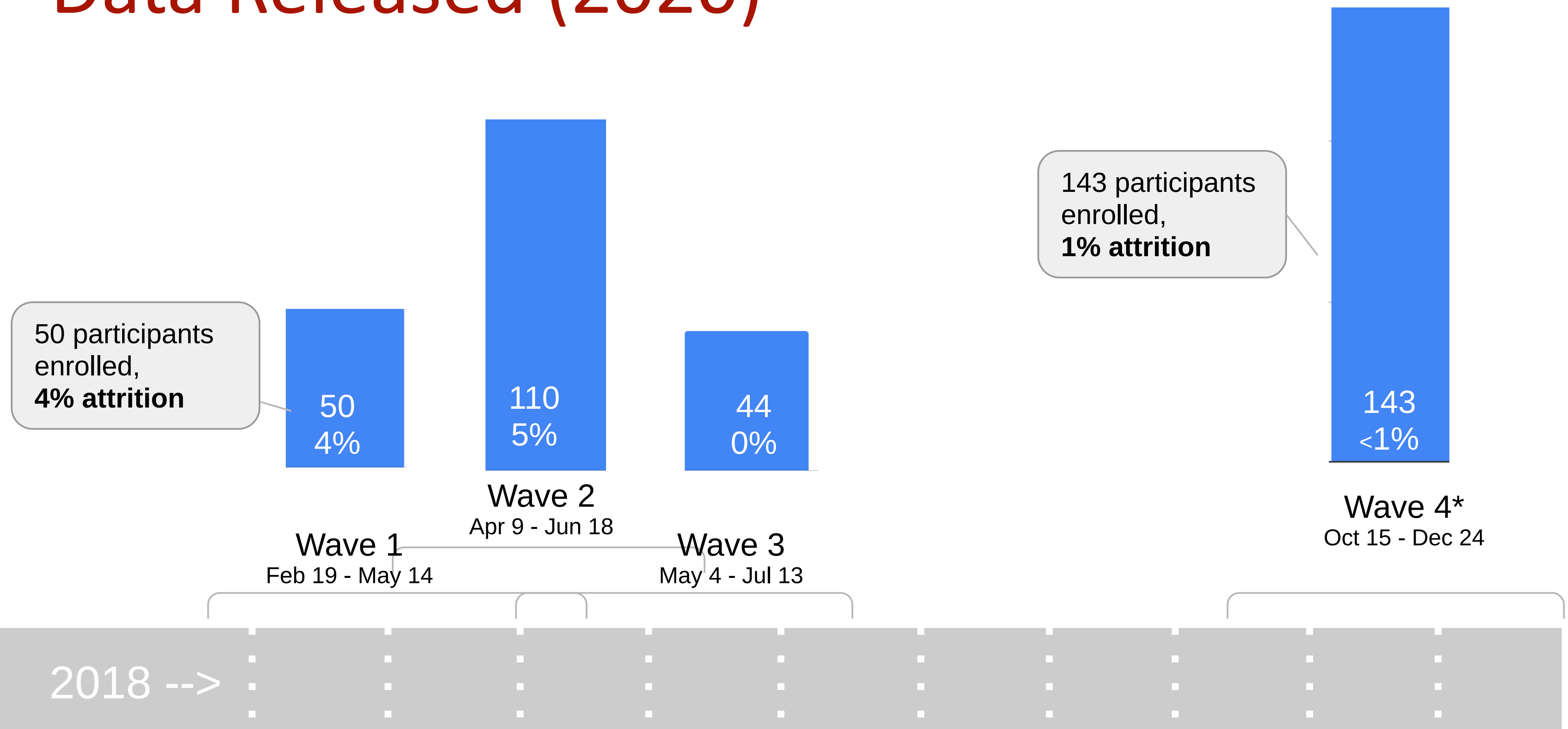


(1) Jelly device	record voice to gather information about pitch, intonation, speed, etc
(2) OMSignal Shirt	track ECG activity & respiratory rate
(3) Fitbit	track physical activity & heart rate
(4) Owl-in-one beacon	map the complex web of movement & social interaction within the hospital & collect various environmental data
(5) RealizD	record how often participants use their phones & for how long
(6) Facebook (7) Instagram	track social media usage via two platforms
(8) TILES App	collect GPS, wifi info, battery level, app analytics & daily surveys
(9) TILES Survey	daily Ecological Moment Assessment survey
(10) MITRE Survey	daily survey from 3rd party — record survey completion & when participants open & close the survey

Study details Phase 1

Keck Medical Center of USC

Data Released (2020)



Waves 1-3
Enrolled 212 (target *N = 200*) out of 375 screened as eligible

Wave 4
Enrolled 144 out of 246 screened as eligible

Karel Mundnich, Brandon Booth, Michelle L'Hommedieu, Tiantian Feng, Benjamin Girault, Justin L'Hommedieu, Mackenzie Wildman, Sophia Skaaden, Amrutha Nadarajan, Jennifer Villatte, Tiago Falk, Kristina Lerman, Emilio Ferrara, and Shrikanth Narayanan. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. Scientific Data (Nature Research). 2020.

Study details Phase 2

PCCM: Pulmonary, Critical Care, and Sleep Medicine

Dates: 11/5/2019-4/13/2020, *Attrition*, N=1 (nurse)

N=8 PCCM Attendings
Interaction Minew Only

N=11 PCCM Fellows
Interaction Minew Only

FOCAL

N=57 Residents
All Sensors and Surveys

N=76 MICU Nurses
Interaction Minew Only

SENSORS


Fitbit Charge 3


Owl-in-One


Atom phone


Rescue time


Minew E8

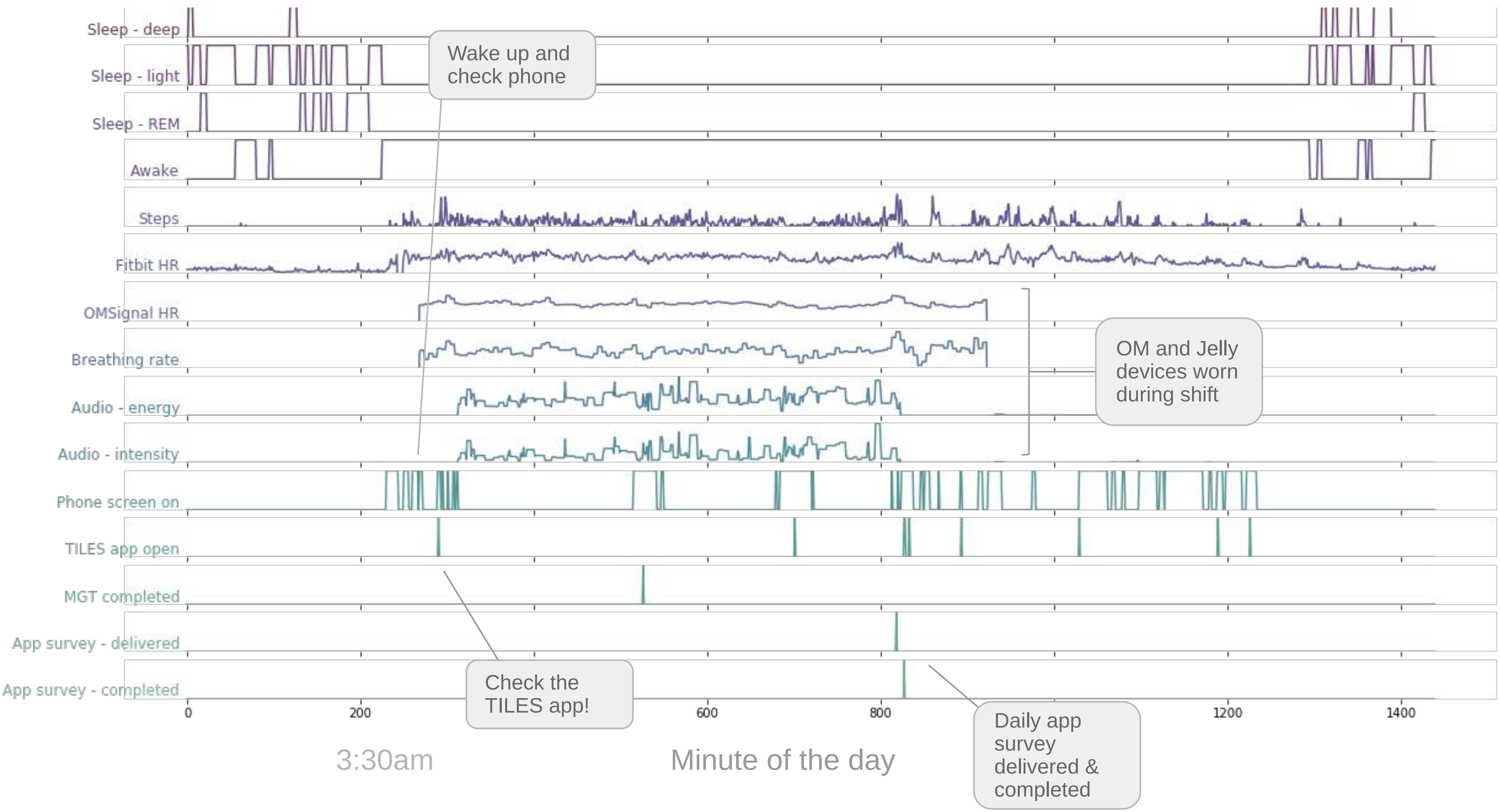
SURVEYS

Initial Prestudy: stress, burnout, activity, sleep, affect, anxiety, depression
Daily: Stress, job load, sleep
Weekly: exercise, stress, team trust, social activity,..

COVID-19 RELATED
Patient Census/details, Equipment, Morbidity, Mortality

A day in the life of a subject

"behavior-gram"



Sleep Differences in Day/Night Shift

DAY SHIFT WORKERS MAINTAIN MORE REGULAR SLEEP PATTERNS

Sleep start

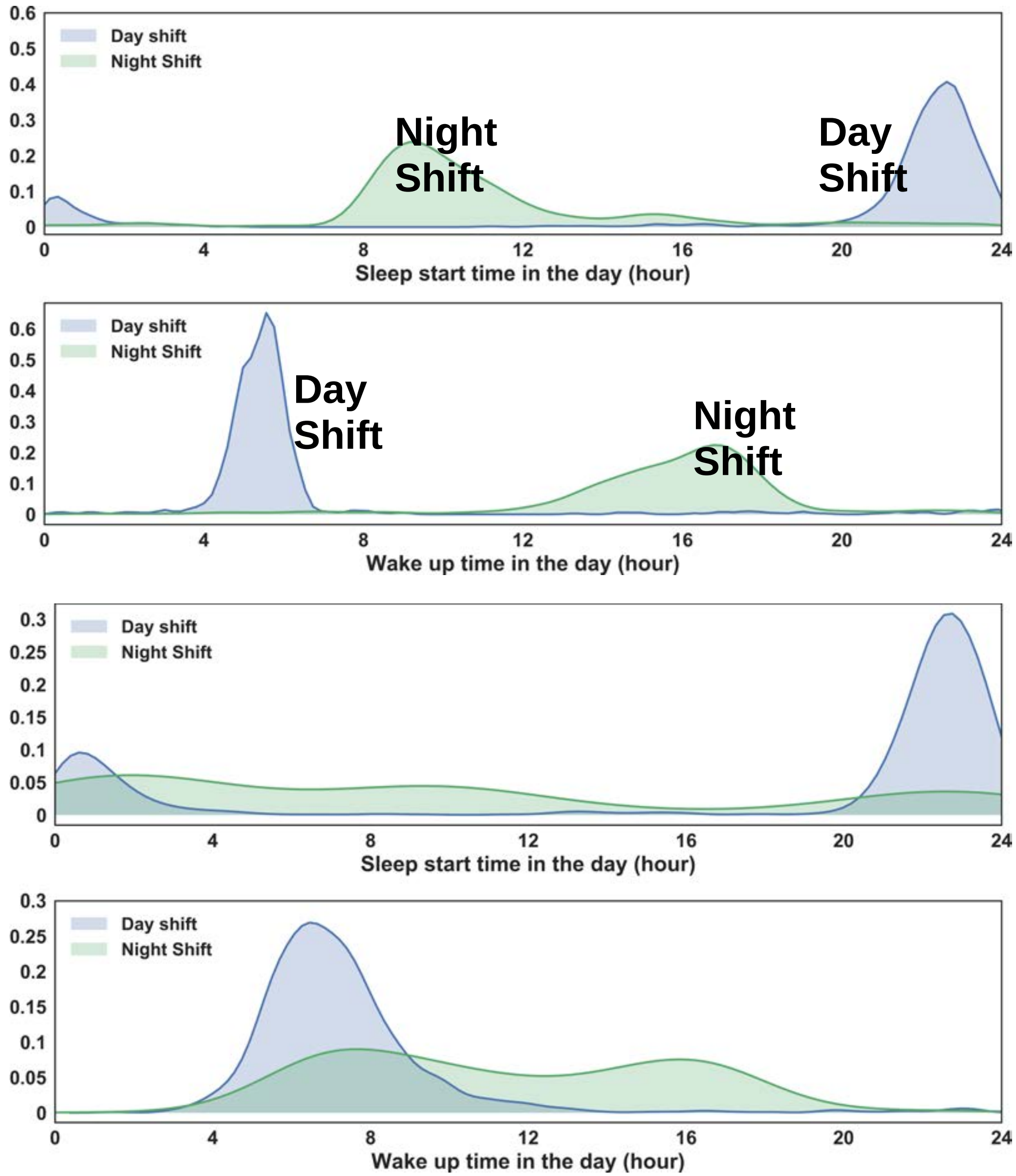
Wake up

WORK DAYS

Sleep start

Wake up

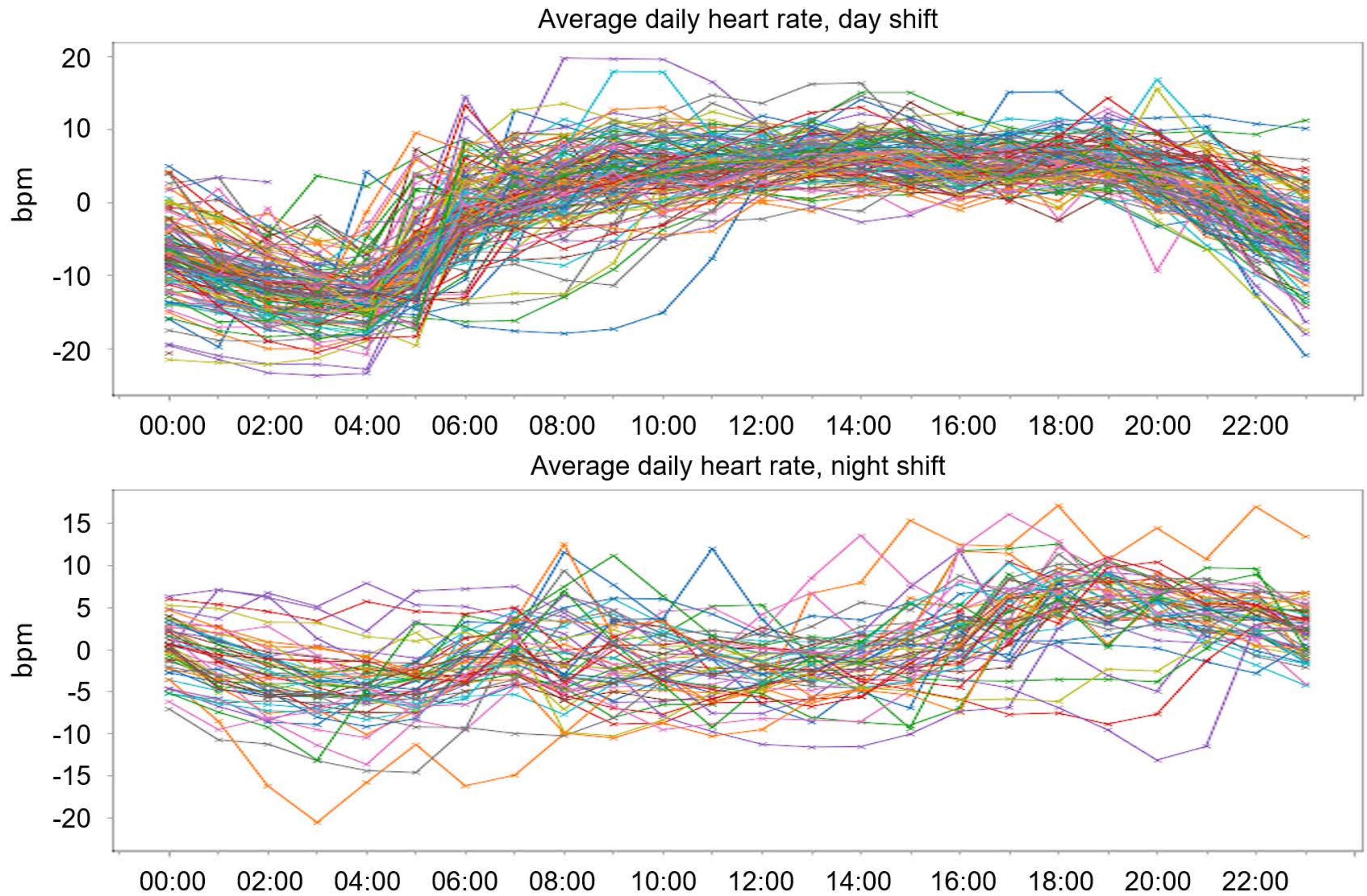
OFF DAYS



Heart Rate Differences in Day/Night Shift

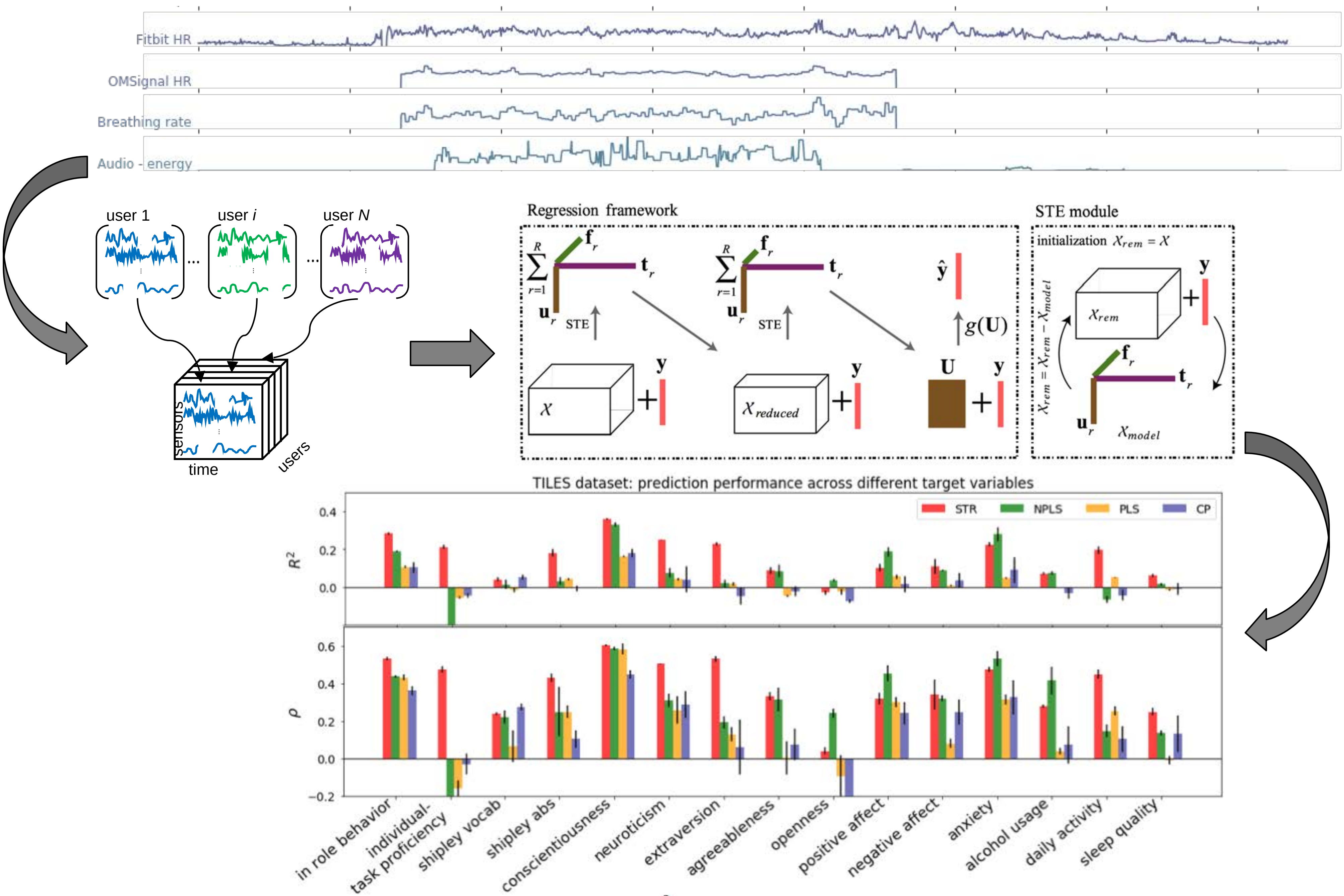
DAY SHIFT WORKERS SHOW MORE REGULAR DAILY HEART RATE PATTERNS

Hourly HR average, mean average in the day is removed. Averaged across days, one line per subject for Day and Night shifts



Tiantian Feng, Brandon M. Booth, Brooke Baldwin-Rodrıguez, Felipe Osorno, Shrikanth Narayanan. A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data. Scientific Reports 11, 8693, 2021

From sensors to models to predictions



References: TILES

1. Tiantian Feng, Brandon M. Booth, Brooke Baldwin-Rodrıguez, Felipe Osorno, Shrikanth Narayanan. A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data. Scientific Reports 11, 8693 (Nature Press). 2021
2. Arindam Jati, Amrutha Nadarajan, Raghuveer Peri, Karel Mundnich, Tiantian Feng, Benjamin Girault, and Shrikanth Narayanan. Temporal Dynamics of Workplace Acoustic Scenes: Egocentric Analysis and Prediction. Proceedings of *IEEE/ACM Transactions on Audio, Speech and Language Processing*. 2021
3. Vinesh Ravuri, Projna Paromita, Karel Mundnich, Amrutha Nadarajan, Brandon M. Booth, Shrikanth S. Narayanan, Theodora Chaspari. Investigating Group-Specific Models of Hospital Workers' Well-Being: Implications for Algorithmic Bias. *International Journal of Semantic Computing*. 2021
4. Amr Gaballah, Abhishek Tiwari, Shrikanth Narayanan, Tiago Falk. Context-Aware Speech Stress Detection in Hospital Workers Using Bi-LSTM Classifiers. Proceedings of ICASSP, Toronto, Canada, May 2021
5. Karel Mundnich, Brandon Booth, Michelle L'Hommedieu, Tiantian Feng, Benjamin Girault, Justin L'Hommedieu, Mackenzie Wildman, Sophia Skaaden, Amrutha Nadarajan, Jennifer Villatte, Tiago Falk, Kristina Lerman, Emilio Ferrara, and Shrikanth Narayanan. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. *Scientific Data* (Nature Research). 2020.
6. Tiantian Feng, Shrikanth Narayanan. Modeling Behavioral Consistency In Large-Scale Wearable Recordings of Human Bio-behavioral Signals. Proceedings of ICASSP, Barcelona, Spain, May 2020
7. Tiantian Feng, Brandon Booth, Shrikanth Narayanan. Modeling Behavior as Mutual Dependency Between Physiological Signals and Indoor Location In Large-Scale Wearable Sensor Study. Proceedings of ICASSP, Barcelona, Spain, May 2020
8. Shen Yan, Homa Hosseinmardi, Hsien-Te Kao, Shrikanth Narayanan, Krisitina Lerman and Emilio Ferrara. Affect Estimation with Wearable Sensors. Journal of Healthcare Informatics Research. 4(3): 261–294, March 2020
9. Michelle L'Hommedieu, Justin H. L'Hommedieu, Cynthia Begay, Alison Schenone, Lida Dimitropoulou, Gayla Margolin, Tiago H. Falk, Emilio Ferrara, Kristina Lerman, and Shrikanth Narayanan. Lessons Learned: Recommendations for Implementing A Longitudinal Study Using Wearable and Environmental Sensors in a Healthcare Organization. J Med Internet Res (JMIR) mHealth and uHealth. 7(12):e13305, December 2019
10. Brandon M Booth, Karel Mundnich, Tiantian Feng, Amrutha Nadarajan, Tiago H Falk, Jennifer L Villatte, Emilio Ferrara, Shrikanth Narayanan. Multimodal Human and Environmental Sensing for Longitudinal Behavioral Studies in Naturalistic Settings: Framework for Sensor Selection, Deployment, and Management. *J Med Internet Res (JMIR),* 21(8):e12832, 2019
11. Abhishek Tiwari, Shrikanth Narayanan, Tiago Falk. Breathing Rate Complexity Features for "In-the-Wild'' Stress and Anxiety Measurement. Proceedings of the 27th European Signal Processing Conference (EUSIPCO), 2019

References: TILES

12. Abhishek Tiwari, Raymundo Cassani, Shrikanth Narayanan, Tiago Falk. A Comparative Study of Stress and Anxiety Prediction in Ecological Settings Using a Smart-shirt and a Smart-bracelet. Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'19), Berlin Germany, July 2019
13. Abhishek Tiwari, Shrikanth Narayanan, Tiago Falk. Stress and Anxiety Measurement "In-the-Wild" Using Quality-aware Multi-scale HRV Features. Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'19), Berlin Germany, July 2019.
14. Tiantian Feng, Shrikanth Narayanan. Imputing Missing Data In Large-Scale Multivariate Biomedical Wearable Recordings Using Bidirectional Recurrent Neural Networks With Temporal Activation. Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'19), Berlin Germany, July 2019
15. Shen Yan, Homa Hosseinmardi, Hsien-Te Kao, Shrikanth Narayanan, Kristina Lerman and Emilio Ferrara. Estimating Individualized Daily Self-Reported Affect with Wearable Sensors. Proceedings of the 7th IEEE Conference on Healthcare Informatics (ICHI2019), Beijing, China, June 2019
16. Amrutha Nadarajan, Krishna Somandepalli, Shrikanth Narayanan. SPEAKER AGNOSTIC FOREGROUND SPEECH DETECTION FROM AUDIO RECORDINGS IN WORKPLACE SETTINGS FROM WEARABLE RECORDERS. Proceedings of ICASSP, Brighton, UK, May 2019
17. Brandon Booth, Tiantian Feng, Abhishek Jangalwa, Shrikanth Narayanan. TOWARD ROBUST INTERPRETABLE HUMAN MOVEMENT PATTERN ANALYSIS IN A WORKPLACE SETTING. Proceedings of ICASSP, Brighton, UK, May 2019
18. Tiantian Feng, Shrikanth Narayanan. Discovering Optimal Variable-length Time Series Motifs in Large-Scale Wearable Recordings of Human Bio-behavioral Signals. Proceedings of ICASSP, Brighton, UK, May 2019
19. Karel Mundnich, Benjamin Girault, Shrikanth Narayanan. Bluetooth based Indoor Localization using Triplet Embeddings. Proceedings of ICASSP, Brighton, UK, May 2019
20. Hsien-Te Kao, Homa Hosseinmardi, Shen Yan, Michelle Hasan, Shrikanth Narayanan, Kristina Lerman and Emilio Ferrara. Discovering Latent Psychological Structures from Self-report Assessments of Hospital Workers. Proceedings of the 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC 2018), Taiwan, November, 2018 [*Best paper award at BESC 2018 for "Distinguished Research on Digital Humanities"*]
21. Tiantian Feng, Amrutha Nadarajan, Colin Vaz, Brandon Booth, and Shrikanth Narayanan. 2018. TILES Audio Recorder: An unobtrusive wearable solution to track audio activity. In WearSys '18: 4th ACM Workshop on Wearable Systems and Applications, June 10, 2018, Munich, Germany.

Some Case Studies

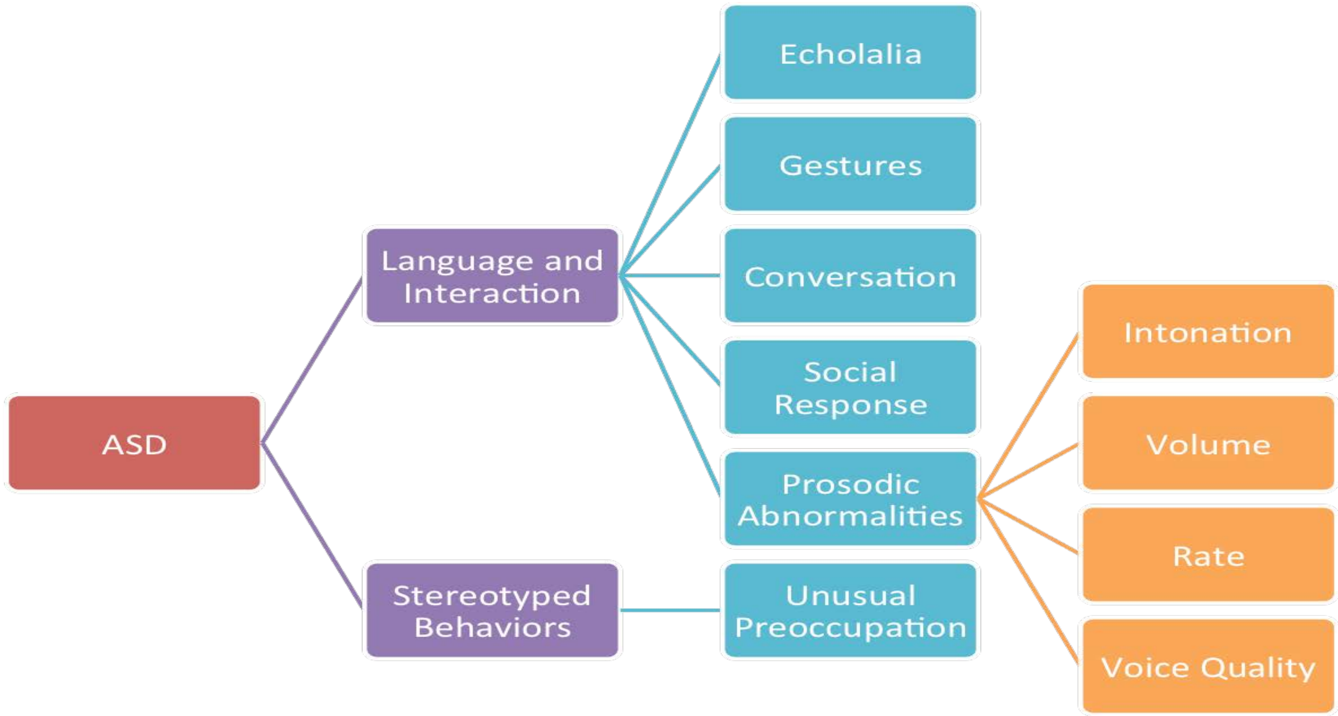
Modeling

Diagnostics

Intervention

Opportunities for rich multimodal approaches in Autism Spectrum Disorder (ASD)

- Better understand communication and social patterns of children
- Stratify behavioral phenotyping with quantifiable and adaptable metrics
- Track, quantify children's progress during interventions



D. Bone, M. Goodwin, M. Black, C-C.Lee, K. Audhkhasi, and S. Narayanan. Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and promises. Journal of Autism and Developmental Disorders. 45(5), 1121-1136, 2015

Daniel Bone, Somer Bishop, Matthew P. Black, Matthew S. Goodwin, Catherine Lord, Shrikanth S. Narayanan. Use of Machine Learning to Improve Autism Screening and Diagnostic Instruments: Effectiveness, Efficiency, and Multi-Instrument Fusion. Journal of Child Psychology and Psychiatry. 57(8): 927-937, August 2016

Quantifying Atypical Prosody

Qualitative descriptions are general and contrasting

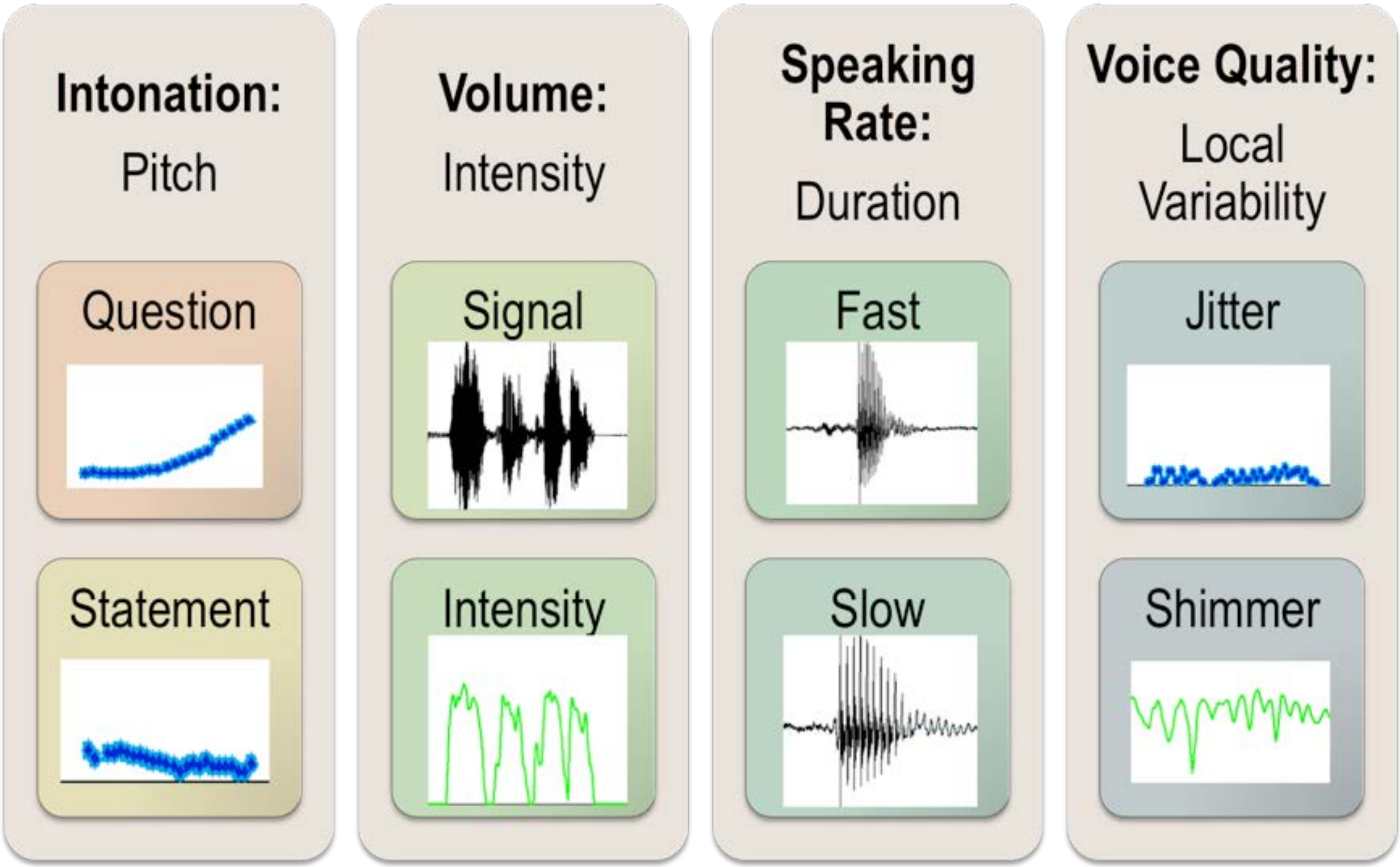
ADOS

Module 3

"slow, rapid, jerky and irregular in rhythm, odd intonation or inappropriate pitch and stress, markedly flat and toneless, or consistently abnormal volume"

Structured assessment may not capture how atypical prosody affects social functioning apart from pragmatics

Quantifying Prosody: Acoustic features



- 24 Features: pitch (6), volume (6), rate (4), and voice quality (8)
  - Intonation: F0 curvature, slope, center
  - Volume: Intensity curvature, slope, center
  - Rate: Boundary (turn end word), Non boundary
  - Voice Quality: Jitter, Shimmer, CPP, HNR

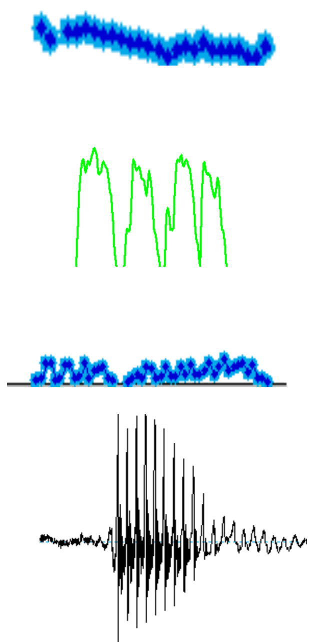
✦ ***median, IQR of above***

Atypical Prosody & Interaction

Spearman's Correlation between rated severity and prosodic cues (dataset ADOS 3 administration, N=28)

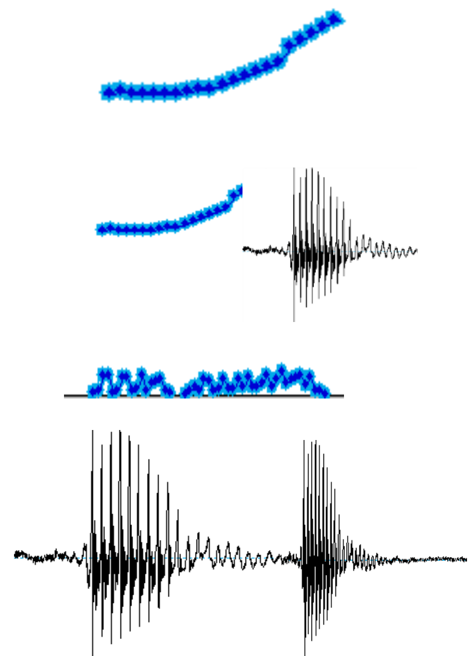
Child's Prosody

- "Monotone"
  $p<0.01$
- "Abnormal volume"
  $p<0.05$
- "Breathy/Rough"
  $p<0.01$
- Slower speaking rate
  $p<0.05$



Psychologist's Prosody

- Questions/affect
  $p<0.05$
- Variable prosody
  $p<0.01$
- also higher jitter
  $p<0.01$
- slower/then faster
  $p<0.01$



The psychologists may be varying their engagement strategies

DANIEL BONE, CHI-CHUN LEE, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT, AND SHRIKANTH NARAYANAN, "THE PSYCHOLOGIST AS AN INTERLOCUTOR IN AUTISM SPECTRUM DISORDER ASSESSMENT: INSIGHTS FROM A STUDY OF SPONTANEOUS PROSODY", JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH, 57:1162–1177, AUGUST 2014.

ASD Severity Regression

Descriptor's Included	Child Prosody	Psych Prosody	Child and Psych Prosody	Underlying Variables
Spearman's ρ	0.50^{**}	0.71^{****}	0.50^{**}	-0.14

*Spearman's ρ between prediction and labels. [**,****]≡a=[0.01,1e-4]. N=28.*

- Multiple linear regression forward-feature selection on the 20 prosodic features, leave-one-session-out

- Psychologist's acoustics more predictive of child's ratings
- Using total feature set shows no advantage.

Modeling Interaction Dynamics Critical

- **More data can offer further insights into prosody, and beyond, in speech communication**

DANIEL BONE, CHI-CHUN LEE, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT, AND SHRIKANTH NARAYANAN, "THE PSYCHOLOGIST AS AN INTERLOCUTOR IN AUTISM SPECTRUM DISORDER ASSESSMENT: INSIGHTS FROM A STUDY OF SPONTANEOUS PROSODY", JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH, 57:1162–1177, AUGUST 2014.

Language and Turn-taking Features

Global Turn-taking Measures (4 features)

- **Can indicate style of interaction**
  - ***speech %***, ***silence %***, ***overlap %*** (*interruption %*), and ***median latency*** (time between turn exchanges)

Rate (3 features)

- **Also useful for characterizing interaction**
  - ***speaking rate*** (*SR*, #-words/utt. dur.; includes pausing)
  - *per-word* **articulation *rate*** (*AR*, syl/word dur.)
  - ***intra-utterance pausing duration***

Language

- **Linguistic Inquiry and Word Count** (LIWC) toolbox
- **Percentages** normalized by the total number of words spoken

* *Examine the whole session, not only the interviews*

(1) words per sentence (WPS)– a rough approximation of mean-length-of-utterance (MLU); (2) first-person, singular pronouns (I, me, mine); (3-5) positive emotion, negative emotion, and affect (positive or negative) language; (6-8) assents (OK, yes), non-fluencies (hm, umm), and fillers (I mean, you know).

DANIEL BONE, CHI-CHUN LEE, THEODORA CHASPARI, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT AND SHRIKANTH NARAYANAN, ACOUSTIC-PROSODIC, TURN-TAKING, AND LANGUAGE CUES IN CHILD-PSYCHOLOGIST INTERACTIONS FOR VARYING SOCIAL DEMAND, INTERSPEECH, 2013.

Results: Language & Turn taking

	Trend w/ Sev.	Psych Feature	Sp. ρ	Trend w/ Sev.	Child Feature	Sp. ρ
Speech Amount	***Increased***	***Speech %***	***0.54***	***Decreased***	***Speech %***	***-0.36***
				Decreased	WPS (MLU)	-0.42
Turn-taking	***Increased***	***Articulatory R***	***0.38***	***Decreased***	***Articulatory R***	***-0.34***
	Increased	Intra-turn sil.	0.32	Increased	Latency	0.34
Pronouns	***Increased***	***Personal Pron.***	***0.38***	***Decreased***	***Personal Pron.***	***-0.40***
Language Use	Decreased	Assent Lang.	-0.48	Decreased	Affect Lang.	-0.40
	Decreased	***Non-fluencies***	***-0.48***	***Decreased***	***Fillers***	***-0.41***

- **Interacting partners influence one another's behavior**
- Child is interacting less with the psychologist
- The psychologist is reacting to child's behavior
- Overall, the conversational quality degrades

- **Psychologist working to engage the child**
- Psychologist may wait for a prompted response, then proceed
- Child may be reluctant to discuss themselves, and may not follow up on prompts to engage psychologist
  - ➢ Child avoid use of the word 'I' [Baltaxe, 1997]
- Psychologist back-channels less, Child uses less fillers

Summary

Objective insights from computational processing

- Prosodic, turn-taking, and language features of the interacting ***psychologist*** and ***child*** indicate the conversational quality degrades for children with greater severity of ASD symptoms
- Psychologist language features may be robust to social demand
- Need for mathematical models of interaction in ASD

Future Work

- Investigate interplay between these varied features
- Larger datasets that include TD and non-ASD DD
- Unsupervised behavioral signals e.g., arousal dynamics, entrainment

DANIEL BONE, CHI-CHUN LEE, THEODORA CHASPARI, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT AND SHRIKANTH NARAYANAN, ACOUSTIC-PROSODIC, TURN-TAKING, AND LANGUAGE CUES IN CHILD-PSYCHOLOGIST INTERACTIONS FOR VARYING SOCIAL DEMAND, INTERSPEECH, 2013.

YOUNG KYUNG KIM, RIMITA LAHIRI, MD NASIR, SO HYUN KIM, SOMER BISHOP, CATHERINE LORD AND SHRIKANTH NARAYANAN. ANALYZING SHORT TERM DYNAMIC SPEECH FEATURES FOR UNDERSTANDING BEHAVIORAL TRAITS OF CHILDREN WITH AUTISM SPECTRUM DISORDER. PROCEEDINGS OF INTERSPEECH, BRNO, CZECH REPUBLIC, 2021

ASD: Understanding the expression of social cues

Production of Affective Facial Expressions (During Smile Imitation Task)

Computational Targets

Quantify atypicality of smile

Region-based activation

Synchrony & symmetry

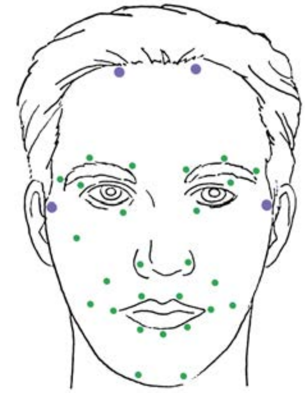
Reduced complexity in dynamic facial behavior primarily in the eye region

- Complexity measured in terms of *Multiscale Sample Entropy (MSE)* [Costa et al. 2011]
- MoCAP data from 20 HFA, 19 TD children, 8 - 12 years of age, no group difference in IQ, age or gender

Tanaya Guha, Zhaojun Yang, Ruth Grossman and Shrikanth Narayanan. A Computational Study of Expressive Facial Dynamics in Children with Autism. IEEE Transactions on Affective Computing. 9(1): 14-20, January 2018

Emily Zane, Zhaojun Yang, Lucia Pozzan, Tanaya Guha, Shrikanth Narayanan, Ruth Grossman. Motion-Capture Patterns of Voluntarily Mimicked Dynamic Facial Expressions in Children and Adolescents With and Without ASD. Journal of Autism and Developmental Disorders. 49(3): 1062-1079, March 2019

Social communication difficulties in autism involve deficits in cross-modal coordination

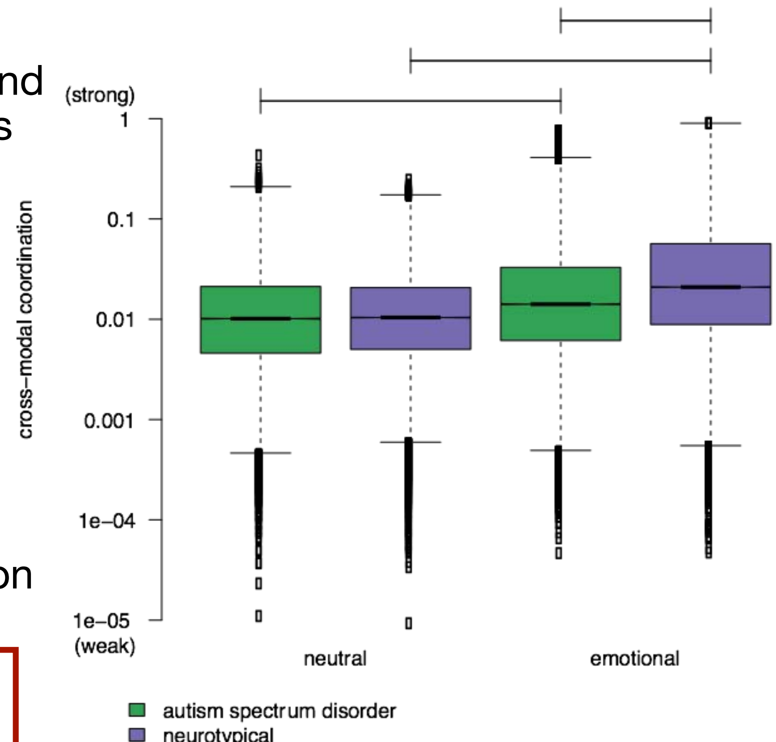


Objective

- Dynamic relation between *speech production and facial expression* in children with autism?
- How face-directed gaze modulates this cross-modal coordination?

Method

- Mimicry task in which participants watched and repeated neutral and emotional spoken sentences with accompanying facial expressions
- Cross-modal coordination measure: Granger causality analysis of dependence between audio and motion capture signals



Results

- Neurotypical children produced emotional sentences with strong cross-modal coordination and produced neutral sentences with weak cross-modal coordination (*differential expressions*)
- Autistic children produced similar levels of cross-modal coordination for both neutral and emotional sentences (*no differentiation*)
- *Cross-modal coordination was greater when the non-ASD child spent more time looking at the face, but weaker when the autistic child spent more time looking at the face*

Tanner Sorensen, Emily Zane, Tiantian Feng, Shrikanth Narayanan, and Ruth Grossman. Cross-Modal Coordination of Face-Directed Gaze and Emotional Speech Production in School-Aged Children and Adolescents with ASD. Scientific Reports (Nature Press). 9, 18301, 2019

Interventions for Addiction

- Motivational Interviewing: Assessment, Training
- Cognitive Behavioral Therapy
- Understanding psychotherapy process mechanisms

USE CASE: "Rate the therapist"—evaluate expressed empathy

B. XIAO, Z. IMEL, P. GEORGIOU, D. ATKINS AND S. NARAYANAN. COMPUTATIONAL ANALYSIS AND SIMULATION OF EMPATHIC BEHAVIORS. A SURVEY OF EMPATHY MODELING WITH BEHAVIORAL SIGNAL PROCESSING FRAMEWORK. CURRENT PSYCHIATRY REPORTS. 2016

DOGAN CAN, REBECA A. MARÍN, PANAYIOTIS GEORGIOU, ZAC IMEL, DAVID ATKINS AND SHRIKANTH NARAYANAN. "IT SOUNDS LIKE...": A NATURAL LANGUAGE PROCESSING APPROACH TO DETECTING COUNSELOR REFLECTIONS IN MOTIVATIONAL INTERVIEWING. JOURNAL OF COUNSELING PSYCHOLOGY. 2015

BO XIAO, ZAC IMEL, PANAYIOTIS GEORGIOU, DAVID ATKINS AND SHRIKANTH NARAYANAN."RATE MY THERAPIST": AUTOMATED DETECTION OF EMPATHY IN DRUG AND ALCOHOL COUNSELING VIA SPEECH AND LANGUAGE PROCESSING. PLOS ONE, 10(12): E0143055. 2015

Modeling Expressed Empathy

- **Speech prosody and empathy: neurological and behavioral evidence of links**
- **Speech prosody measures : turn duration, energy, pitch, jitter, shimmer**

MI based psychotherapy

Human
observation
& annotation

Audio recording → Denoising → Utterance segmentation → Prosodic features → Feature quantization → Distribution of prosodic patterns → Therapist Empathy ratings

Automatic Inference

- Perceived empathy measured by prosodic behavioral cues
- Joint model of both speakers

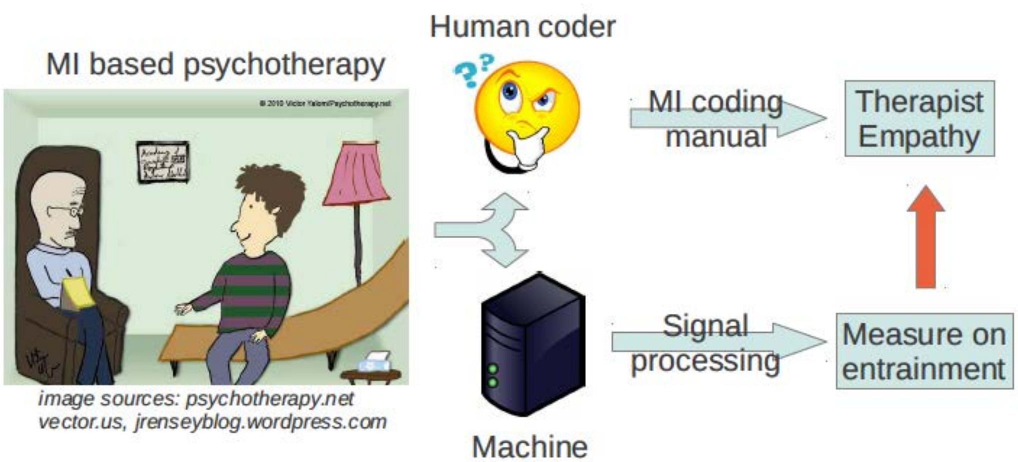
Key Findings

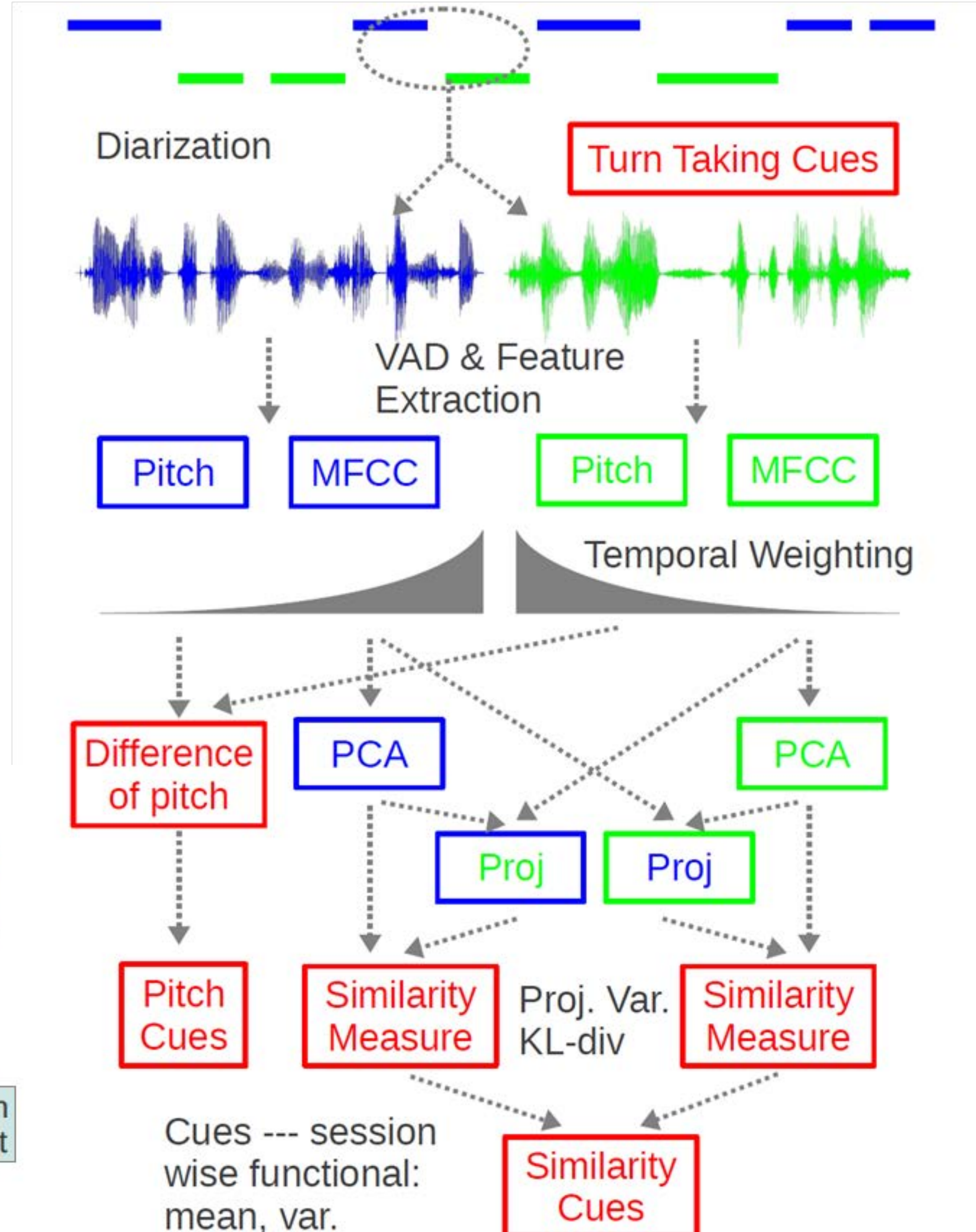
- **lower perceived empathy of therapist when:**
  - Therapist has higher energy values
  - Therapist has higher pitch values

Bo Xiao, Daniel Bone, Maarten Van Segbroeck, Zac E. Imel, David Atkins, Panayiotis Georgiou and Shrikanth Narayanan, Modeling Therapist Empathy through Prosody in Drug Addiction Counseling, in: Proceedings of Interspeech, 2014

Vocal Entrainment Measures

- **Link between entrainment measures and perceived empathy**
  - Behavior of interlocutors become similar
  - Define similarity metrics on speech-derived properties
  - **Found significant correlation: higher entrainment/similarity implies higher empathy**



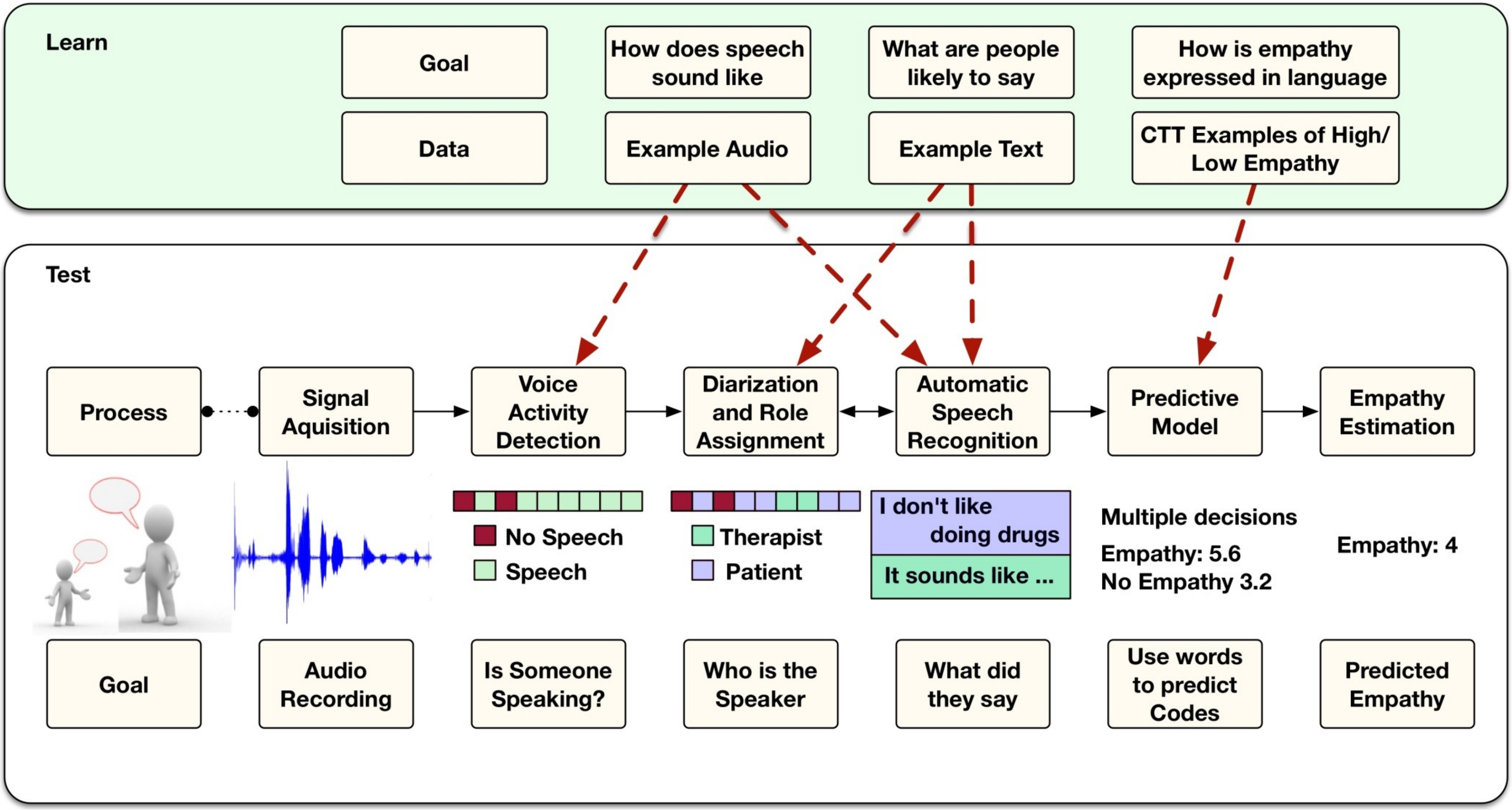


Bo Xiao, et al., Modeling Therapist Empathy and Vocal Entrainment in Drug Addiction Counseling Proceedings of Interspeech, 2013

“Sound to code” system:

Used in mental health clinics

Estimating empathic behavior directly from audio



– **82%** accuracy for *fully* automatic system (no human intervention)

– **61%** (chance), **85%** (manual transcripts), **90%** (human agreement)

Bo Xiao, Zac Imel, Panayiotis Georgiou, David Atkins and Shrikanth Narayanan."Rate my therapist": Automated detection of empathy in drug and alcohol counseling via speech and language processing. PLoS ONE, 10(12): e0143055. 2015

Multi-label Multi-task Modeling: Psychotherapy Behaviors

Generalizing

across domains: Motivational Interviewing, Cognitive Behavioral Therapy,…

- **Multi-label learning**
  - benefits prediction of less frequently occurring behaviors by leveraging modeling of more frequent behaviors
- **Multi-task learning**
  - benefits prediction of behaviors across domains by modeling common behaviors
- **Modeling user-turn context useful**
- **Evaluation on two psychotherapy approaches**
  - *Motivational Interviewing* (11 aggregate MISC codes; 345 sessions)
  - *Cognitive Behavioral Therapy* (11 CTRS codes; 92 sessions)
  - Deep Multi label Multi task Context aware learning: >5% absolute improvement in code prediction for both domains

J. Gibson, D. Atkins, T. Creed, Z. Imel, P. Georgiou and S. Narayanan, "Multi-label Multi-task Deep Learning for Behavioral Coding," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2952113. 2019

N. Flemotomos, V. Martinez, Z. Chen, K. Singla, V. Ardulov, R. Peri, D. Caperton, J. Gibson, M. Tanana, P. Georgiou, J. Van Epps, S.. Lord, T. Hirsch, Z. Imel, D. Atkins, and S. Narayanan. Automated Evaluation Of Psychotherapy Skills Using Speech And Language Technologies. Behavior Research Methods. 2021

Computational Media Intelligence

Special focus on Diversity and Inclusion

- understanding media stories, and their impact on human experiences, behavior and action: from individual to socio-cultural scale
- support diversity and inclusion: tools for awareness, tools for change

Case study: Quantifying Media Portrayals

- **Understand gender, age, race representations**
  - on screen *and* behind the scenes

- **But can go beyond measuring (unconscious) bias and stereotypes..**
  - Provide insights into positive, societally meaningful portrayals e.g., of STEM
  - Assist creators with analytical tools during the creative process
  - Enable quantitative causal models for decision making

In collaboration with

Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

With support from

Google

Illustration: On-Screen Time, Speaking Time

Screen-time

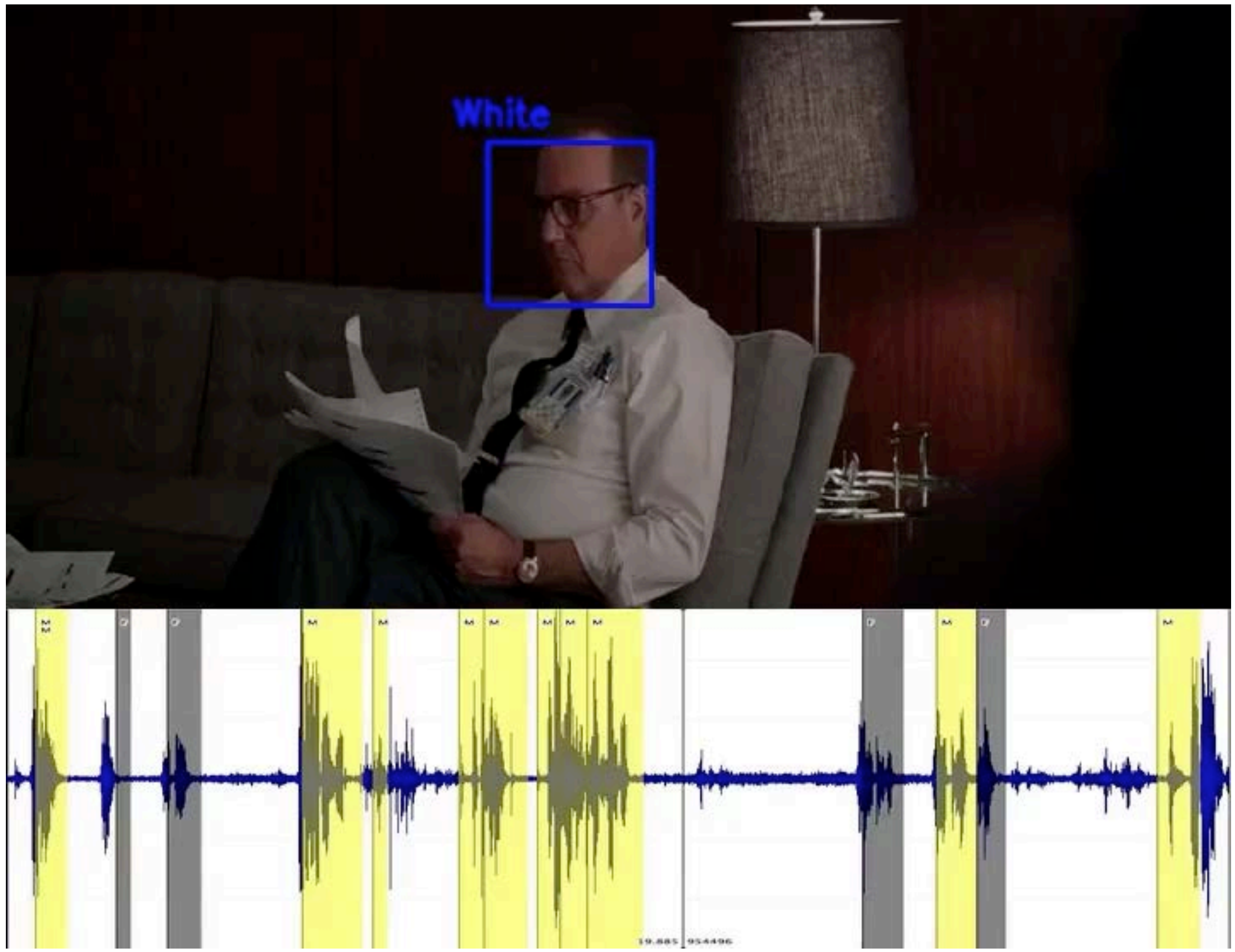
Female face

Male face

Speaking-time

Female speaker

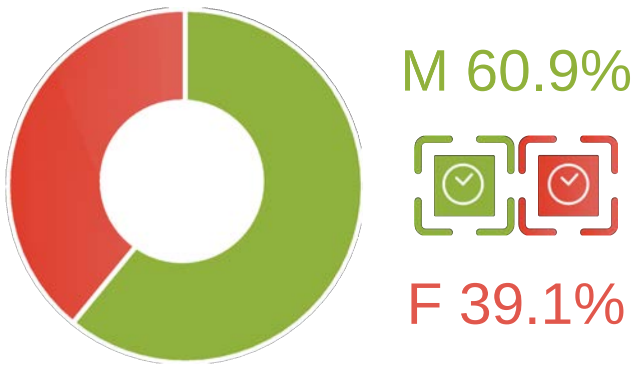
Male speaker



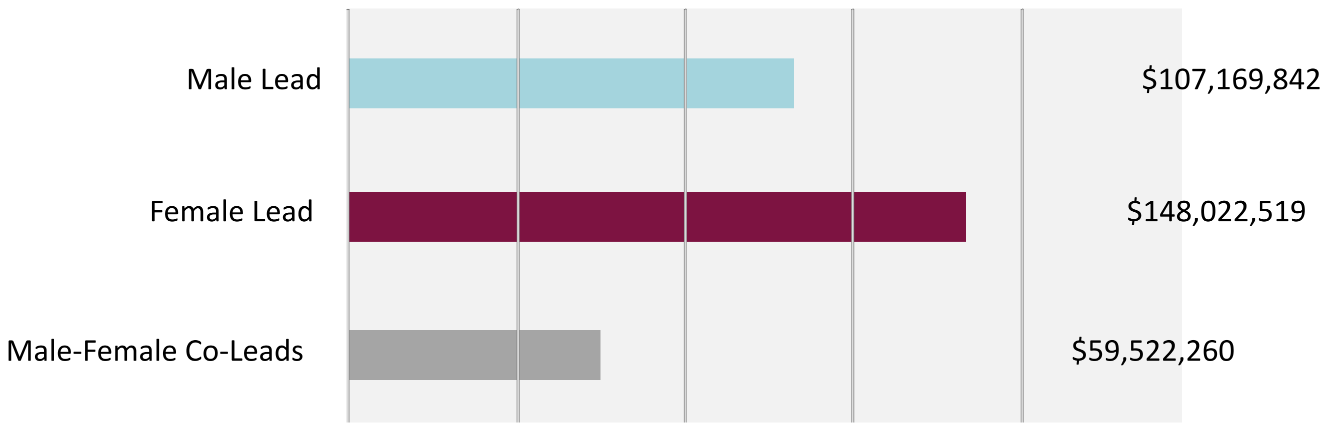
T. Guha, C.-W. Huang, N. Kumar, Y. Zhu, and S. S. Narayanan, "Gender representation in cinematic content: A multimodal approach," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015, pp. 31–34.

On top grossing ~100 live action US Films for 2017, 2018

2017 Screen Time by Gender (N = 94)



2017 Speaking Time by Gender (N = 94)



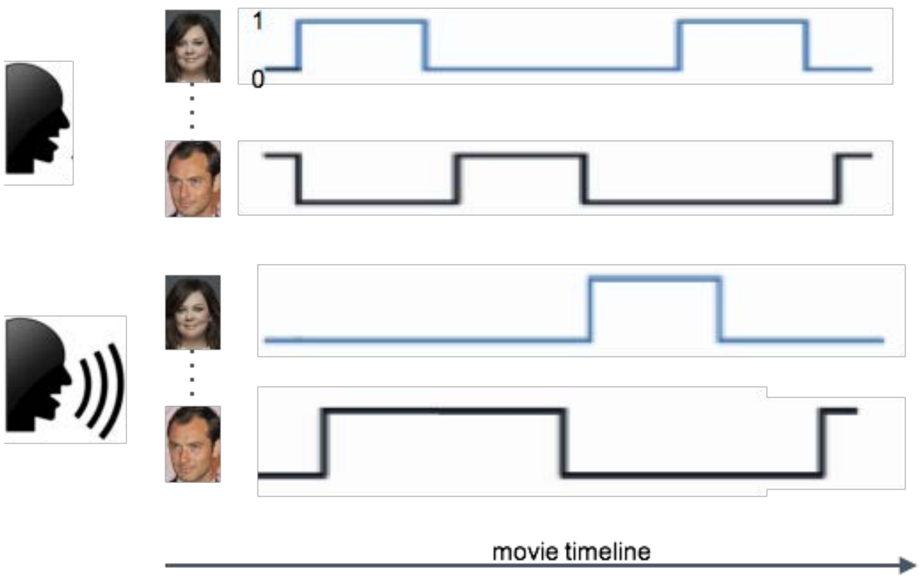
November 7, 2019: Analysis of 2.7 Million Ads—30% more views for ads with gender parity

Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

Google

Joint Audio-visual Analysis: Sample insights

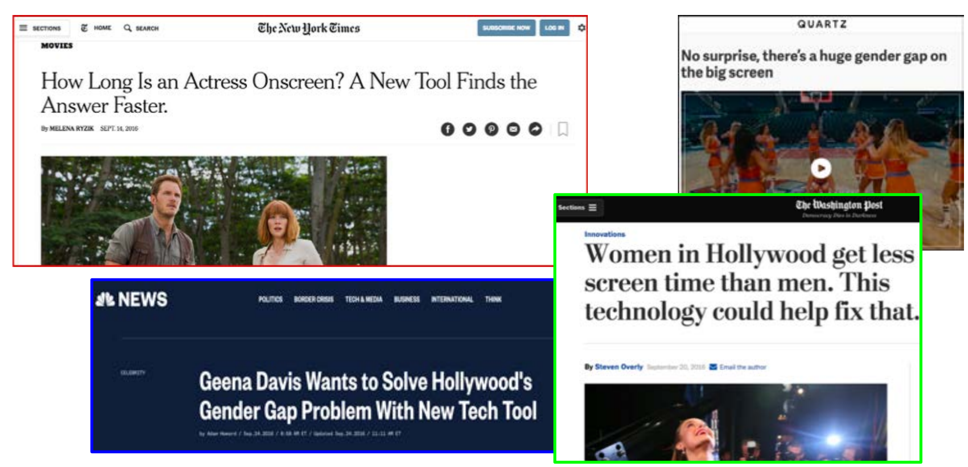
representational disparity



Data from 17 Hollywood blockbusters..

	No face	male face	female face
No speech	26.5%	49.7%	24.8%
male voice	20.9%	51.1%	28.0%
female voice	16.6%	50.4%	33.0%

.... seen less even while speaking

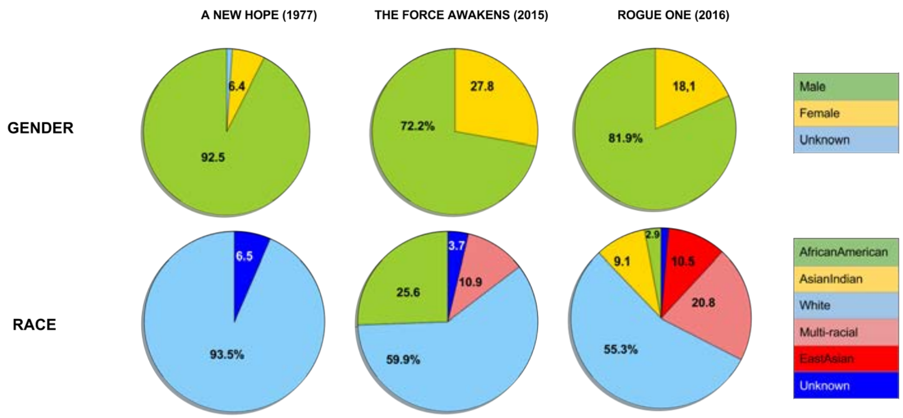


Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

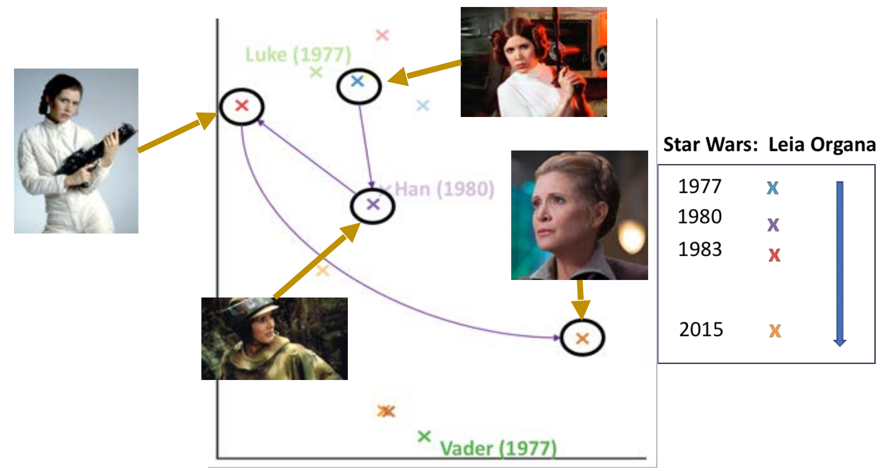
Text Analytics and Natural Language Processing

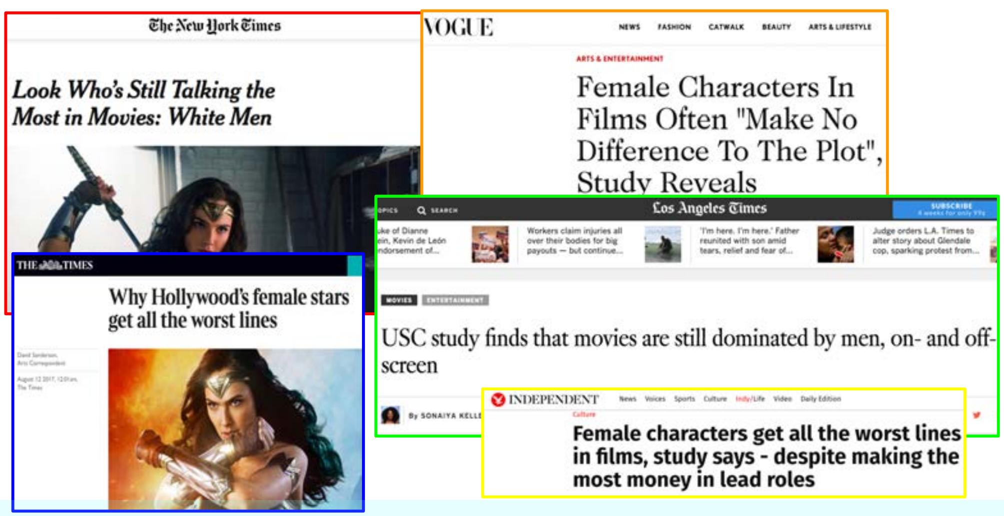
Dialog and interaction language analytics from text documents
e.g., scripts, books, subtitles: *who is saying what to whom and how*

Representations over time: A case study of Star Wars trilogy



Distribution of dialogues for race and gender from movie scripts



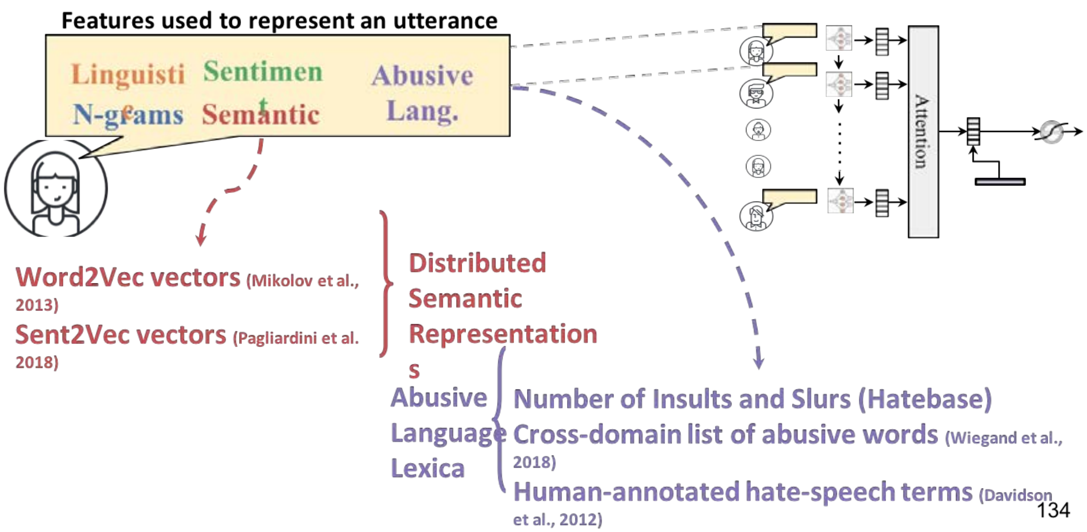


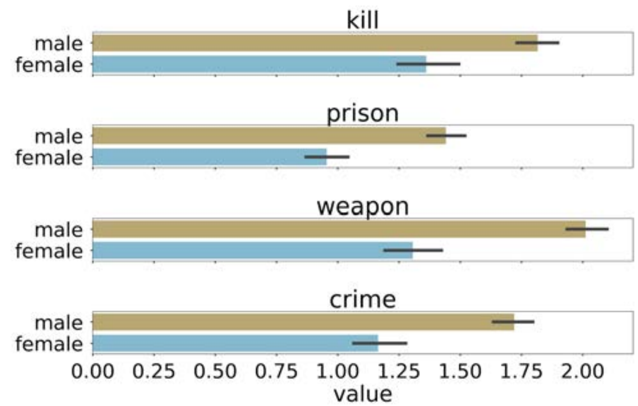
Ramakrishna, A., Malandrakis, N., Staruk, E., & Narayanan, S. (2015). A quantitative analysis of gender differences in movies using psycholinguistic normatives. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 1996–2001).

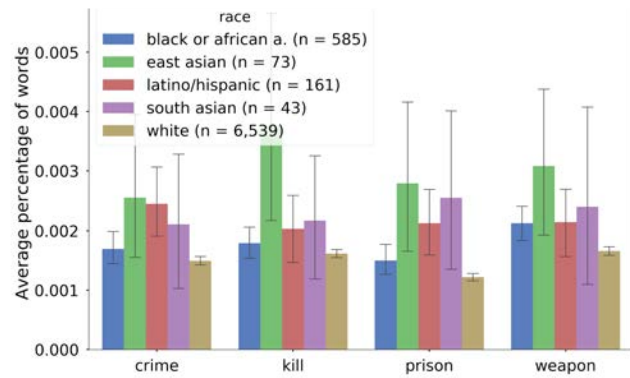
Risk Behaviors in movies

Exposure to violent, sexual, or substance-abuse content in media increases children's and adolescents' willingness to imitate similar behaviors.

Violent Content in Movies: Results Snapshot







Data: Movie Screenplay Dataset (Ramakrishna et al., 2017)
954 Hollywood movie scripts (1920-2016)
12 different genres, Main characters identified

- Male characters use more explicit abusive language than female characters
- Significant differences between racial groups
  - White characters use the least
  - East Asians use the most
- Interactions between male characters more abusive than any other type of interaction

Victor Martinez, Krishna Somandepalli, Karan Singla, Anil Ramakrishna, Yalda Uhls, Shrikanth Narayanan. Violence Rating Prediction from Movie Scripts. Proceedings of Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019
Anil Ramakrishna, Victor R. Martínez, Nikolaos Malandrakis, Karan Singla and Shrikanth Narayanan. Linguistic analysis of differences in portrayal of movie characters. Proceedings of the 55th ACL, 2017

Joint Estimation and Analysis of Risk Behavior Ratings in Movies

Risk Behavior Portrayals:

- Tend to co-occur with one another (Bleakley et al., 2014, 2017; Thompson & Yokota, 2004)
- Potential side effects of repeated exposure esp. children and teens (Anderson & Bushman, 2001)
- *Current Solutions:* Depend on audio-visual cues, only available in post-production
- *Our solution:* Estimate risk behavior content ratings from language used in movie scripts

Automatic identification of Risk Behaviors in Movie scripts:

- Multi-task learning to capture co-occurrence of violence, sexual and substance abuse
- Sequence learning on the semantics and affective aspects of what characters say
- Model *significantly outperforms* SoTA for violent content classification (Martinez et al., 2019)

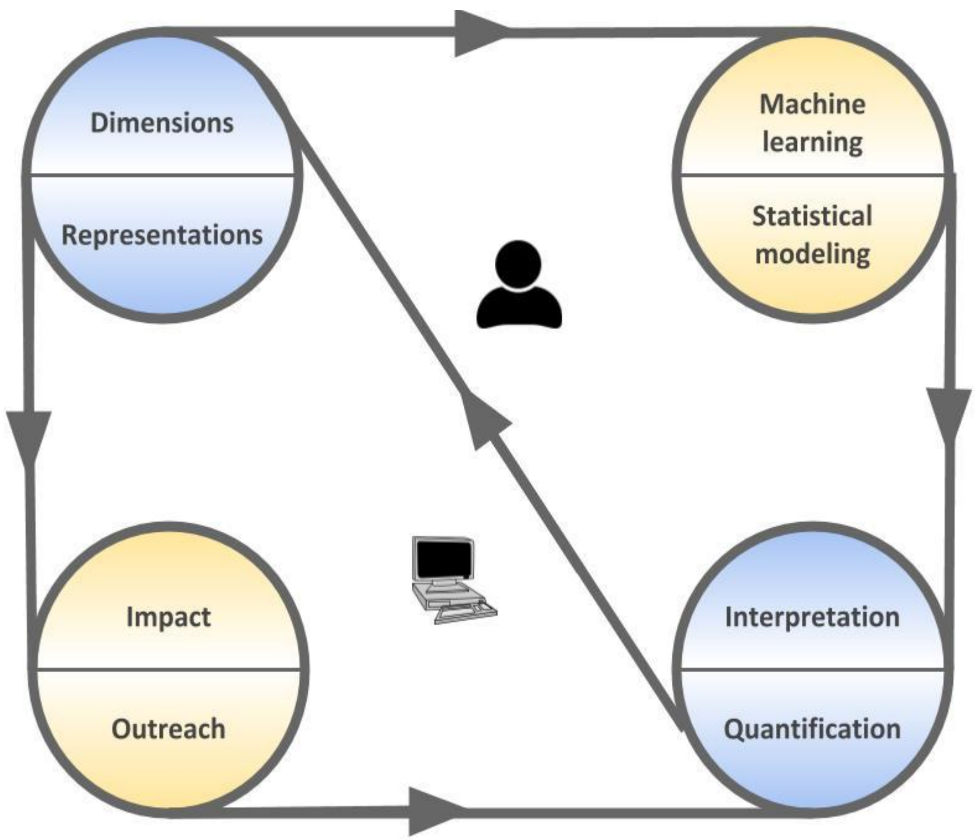
Highlights:

- Film-makers compensate low levels of violent content with both sexual and substance abuse content
- MPAA ratings are much more sensitive towards sexual content than to violent content

Victor Martinez, Krishna Somandepalli, Yalda Tehranian-Uhls and Shrikanth Narayanan. Joint Estimation and Analysis of Risk Behavior Ratings in Movie Scripts. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

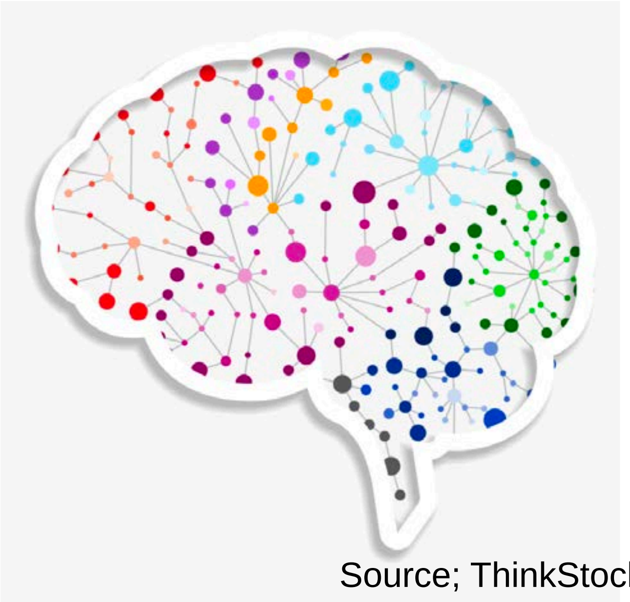
USC Center on Computational Media Intelligence

Create and deploy engineering systems at scale to understand media content, in its multiple modalities, over time, and measure its impact on individuals and society



Help Fill Gaps

Help Connect Dots

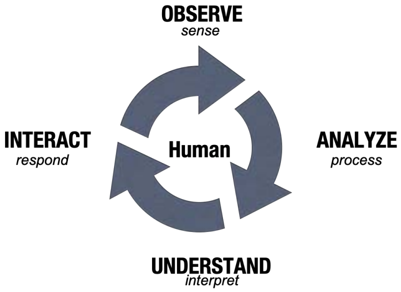


Source; ThinkStock

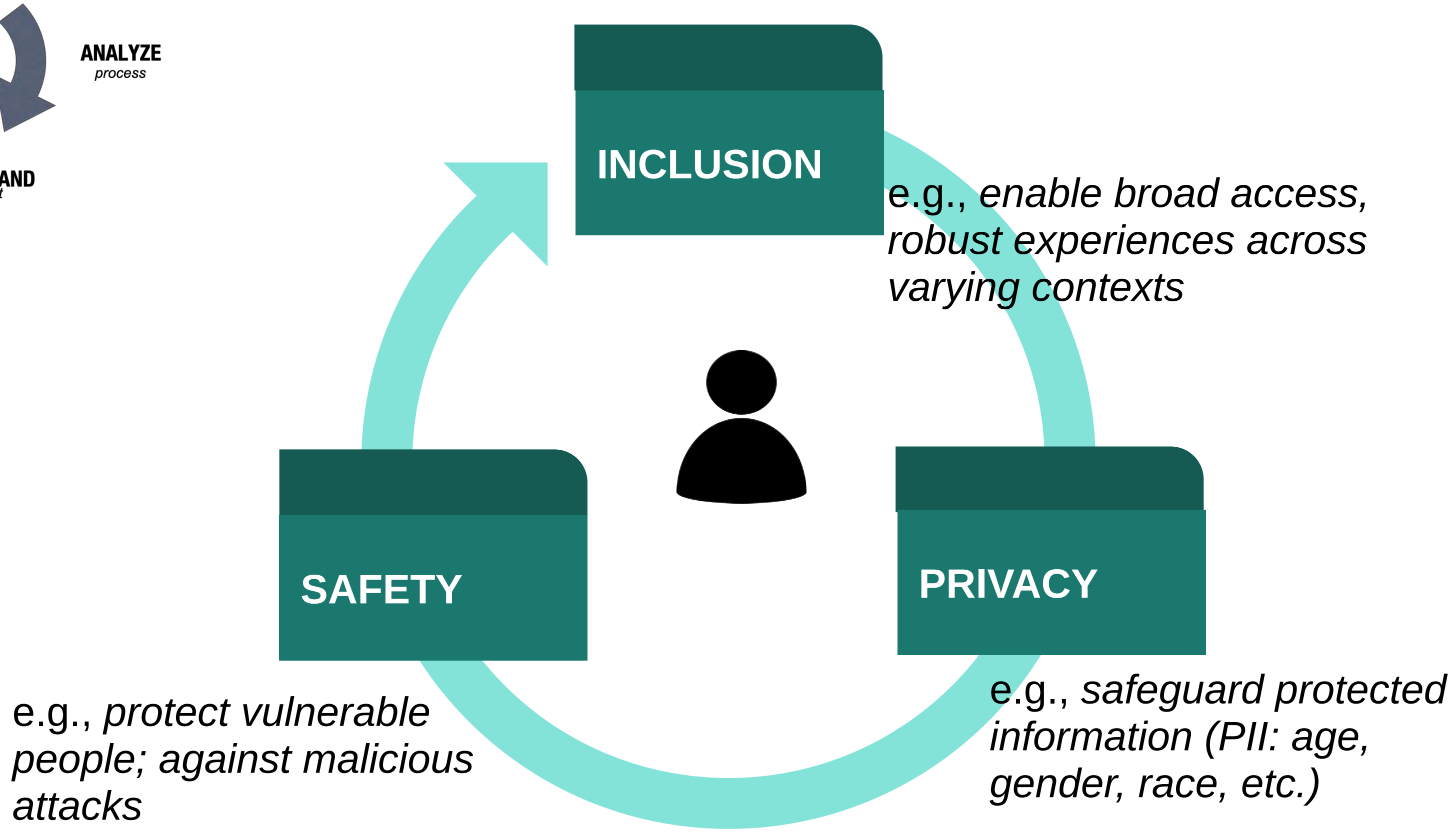
Open Challenges → Rich Opportunities

- **Getting the right multimodal data**
  - *sensing* in natural settings; capturing context
  - doing it in a time "sensitive" way
- **Processing the data**
  - variability, heterogeneity and uncertainty in data
  - specifying behavior representations for computing
  - reflecting multiple (diverse) perspectives & subjectivity
  - interpretable, targetable "features" for interventions
  - dealing with various levels of "imperfect" solutions
  - learning/transfer across domains
- **Using the data, closing the loop with stakeholders**
  - Data provenance, privacy, trust, integrity, sharing
  - Enabling interventions & evaluation at scale, cost, JIT
  - Choosing the right operating point: adaptivity

Inclusive technologies key ingredient of enabling
Trustworthy Human-centered Machine Intelligence



Speech is critical part of this ecosystem



others
explainability
transparency
resilience
…

96

Twin goals: Understanding and addressing variability within and across people and their contexts

S. Narayanan and A. Madni. Inclusive Human centered Machine Intelligence. *The Bridge*. 50(S): 113-116. National Academy of Engineering, 2020

Shared Multimodal Resources: Critical

From

Databases

IEMOCAP Database
The Interactive Emotional Dyadic Motion Capture (IEMOCAP) database is an acted, multimodal and multispeaker database, recently collected at SAIL lab at USC. It contains approximately 12 hours of audiovisual data, including video, speech, motion capture of face, text transcriptions. (Read more...)

MICA Text Corpus
The MICA Text Corpus is now available for download. (Read more...)

EMA Database
The Electromagnetic Articulography (EMA) database contains a total of 680 utterances spoken in four different target emotions, happiness, sadness and neutrality. (Read more...)

MRI-TIMIT Database
MRI-TIMIT is a large-scale database of synchronized audio and real-time magnetic resonance imaging (rtMRI) data for speech res base currently consists of midsagittal upper airway MRI data and phonetically-transcribed companion audio, acquired from two female speakers of American English. (Read more...)

USC-TIMIT Database
USC-TIMIT is a database of speech production data under ongoing development, which currently includes real-time magnetic re data from five male and five female speakers of American English, and electromagnetic articulography data from three of these s more...)

CreativeIT Database
The CreativeIT database is an acted and multimodal database of dyadic theatrical improvisations. It contains 8 sessions of audio cluding video, speech, and full-body motion capture data. (Read more...)

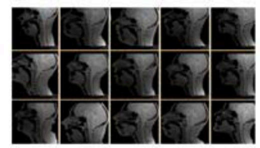
VAM Database
The Vera am Mittag (VAM) database is a German audio-visual speech database recorded from a talk-show on TV. Its main corpus most 1,000 labelled audio samples of spontaneous,unscripted emotional expressions. (Read more...)

https://sail.usc.edu/software/databases/

resources

Please click on the images below for access to research and educational resources that the SPAN group makes publicly available.

speech morphology database

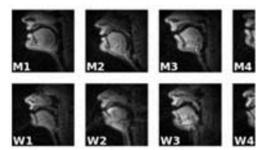


the USC Speech and Vocal Tract Morphology MRI Database includes real-time and volumetric MRI from 17 speakers

the rtMRI IPA charts

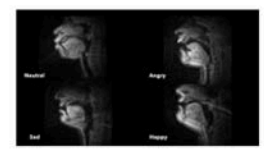
watch real-time MRI videos corresponding to the sounds of the International Phonetic Alphabet

USC-TIMIT



USC-TIMIT is a real-time MRI and electromagnetic articulography database with companion software tools freely available for research purposes

USC-EMO-MRI



a real-time MRI database capturing speech production across emotions

- Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette Chang, Sungbok Lee, and Shrikanth Narayanan. IEMOCAP: Interactive emotional dyadic motion capture database. Journal of Language Resources and Evaluation. 42:335-359, 2008.
- Michael Grimm, K. Kroschel, and S. Narayanan. The Vera Am Mittag German Audio-Visual Emotional Speech Database. In Proc. International Conference on Multimedia and Expo, 2008.
- Angeliki Metallinou, Zhaojun Yang, Chi-Chun Lee, Carlos Busso, S. Carnicke and S. Narayanan. The USC CreativeIT Database of Multimodal Dyadic Interactions: From Speech and Full Body Motion Capture to Continuous Emotional Annotations. Journal of Language Resources and Evaluation. pp. 1-25, 2015
- … …
- Karel Mundnich, Brandon Booth, M. L'Hommedieu, T. Feng, B. Girault, J. L'Hommedieu, M. Wildman, S. Skaaden, A. Nadarajan, J. Villatte, T. Falk, K. Lerman, E. Ferrara, and S. Narayanan. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. Scientific Data (Nature Research). 2020.

HUMAN-CENTERED MACHINE INTELLIGENCE:

SUPPORT HUMAN &/OR AUTONOMOUS DECISION MAKING, ACTION & RESPONSE USING SENSING, DATA SCIENCES AND AI TECHNOLOGIES

✓HELP US DO THINGS WE KNOW TO DO MORE EFFICIENTLY, CONSISTENTLY

➡MODEL AND PREDICT CONSTRUCTS E.G., EMOTIONS, ENGAGEMENT

✓HELP HANDLE NEW DATA, CREATE NEW MODELS TO OFFER NEW INSIGHTS

➡CREATE TOOLS FOR SCIENTIFIC DISCOVERY E.G., AFFECT REGULATION

✓HELP CREATE TOOLS TO SUPPORT DIAGNOSTICS, PERSONALIZED INTERVENTIONS, AND TRACKING RESPONSE TO TREATMENT

- Shrikanth Narayanan and Panayiotis Georgiou. Behavioral Signal Processing: Deriving Human Behavioral Informatics from Speech and Language. Proceedings of IEEE. 101(5): 1203-1233, May 2013
- Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications. IEEE Signal Processing Magazine. 34(5): 189-196, September 2017
- Krishna Somandepalli, Tanaya Guha, Victor Martinez, Naveen Kumar, Hartwig Adam, Shrikanth Narayanan. Computational Media Intelligence: Human-centered Machine Analysis of Media. Proceedings of IEEE. 2021

Work reported represents efforts of numerous colleagues and collaborators

Too many to name, but grateful to all

SUPPORTED BY:
NSF, NIH, ONR, ARMY, DARPA, IARPA, SIMONS FOUNDATION

Signal Analysis and Interpretation Laboratory

....technologies to understand the human condition
and to support and enhance human capabilities and experiences

creating inclusive technologies and technologies for inclusion

Shri Narayanan, Director

http://sail.usc.edu