Human-centered Machine Intelligence

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

University of Southern California

TAPAS-June 2020

Signal Analysis and Interpretation Laboratory

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

SAiL

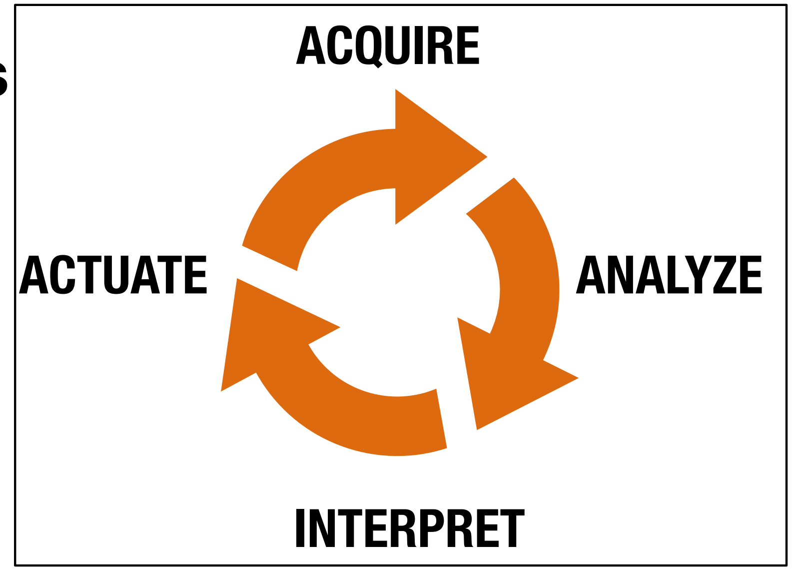
creating inclusive technologies and technologies for inclusion

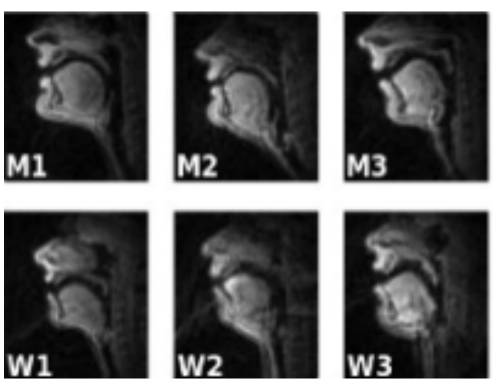
USC
School of Engineering

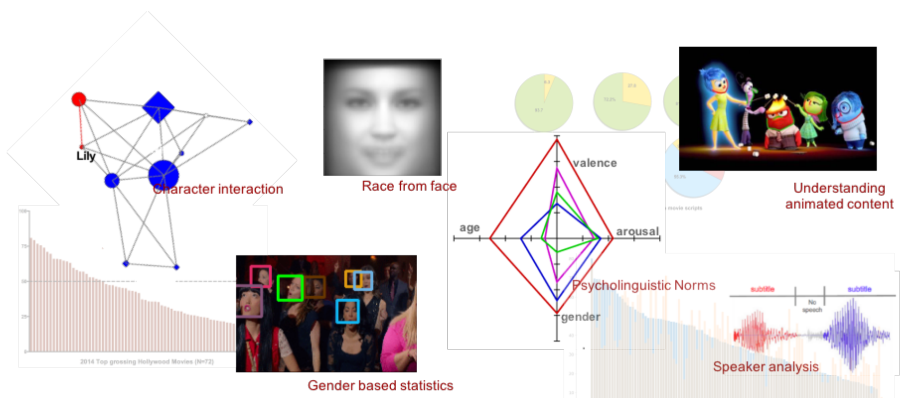
http://sail.usc.edu

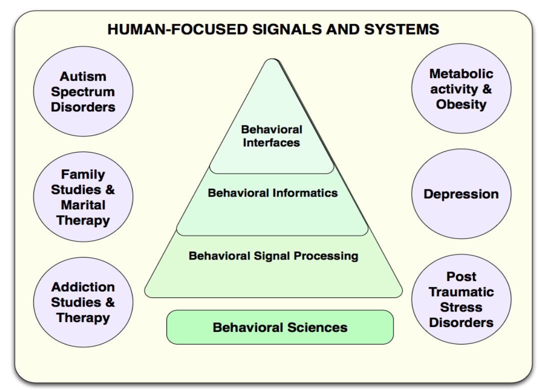
Highlights from three areas of Human-centered Machine Intelligence

- Speech Science and Clinical Applications
- Computational Media Intelligence
- Behavioral Machine Intelligence









3

Highlight 1

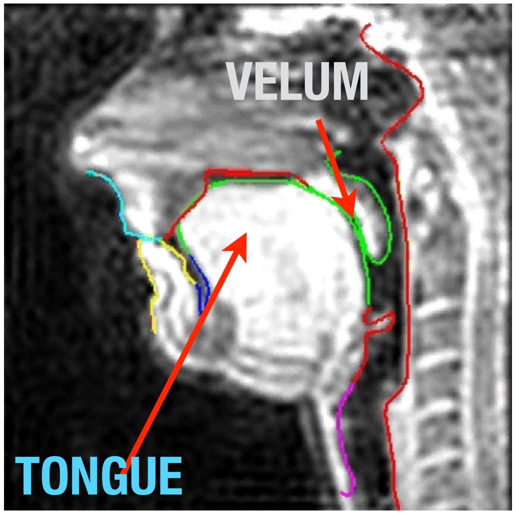
Computational Human Communication Science

From real time MRI to scientific discovery and clinical advances

- investigating speech and language production: from its cognitive conception, to its bio-mechanical execution, to its signal properties
- technology applications in speech recognition, biometrics, synthesis
- diagnostic and therapeutic applications in cancer, neurological disorders

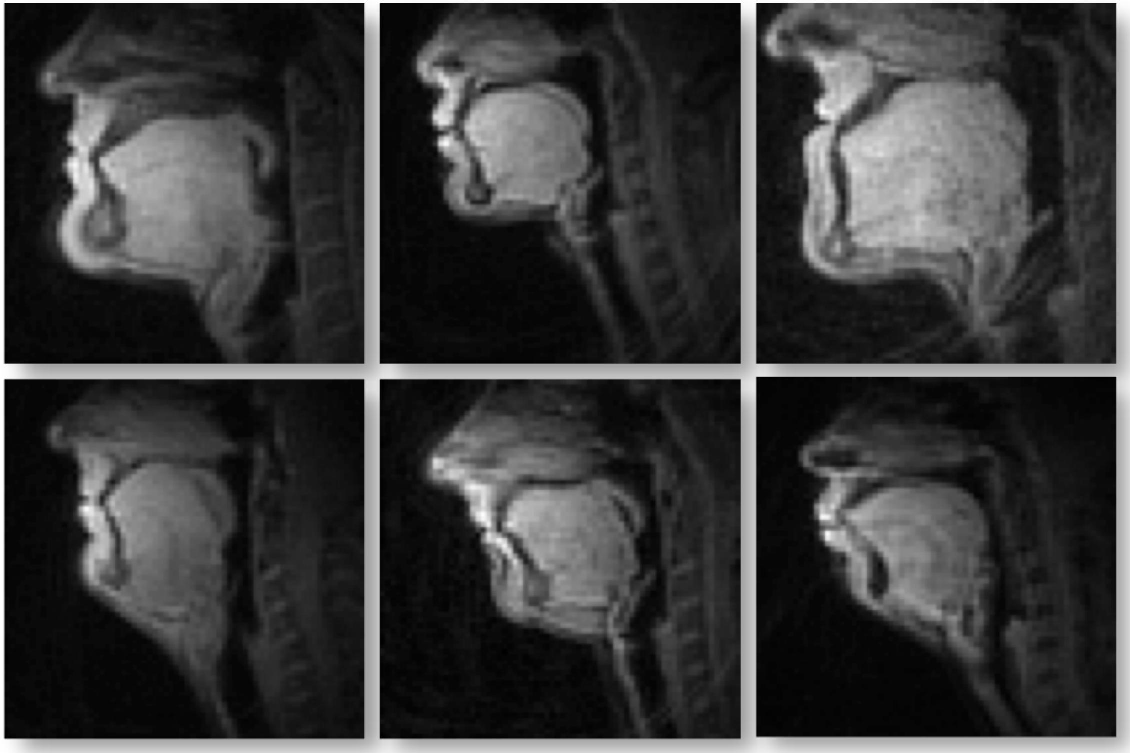
First to create Real-time Magnetic Resonance Imaging System to see the human vocal instrument in action

24 frames/s: 2008

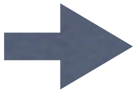


>100 frames/s: 2018

Different individuals….



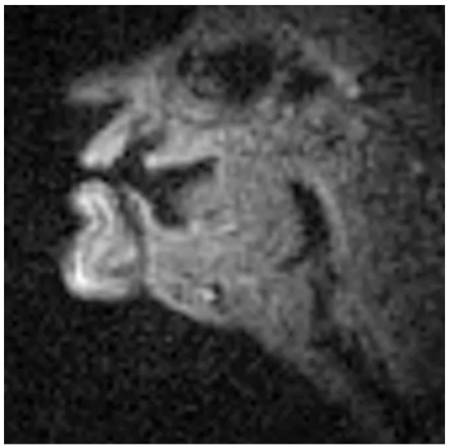
..each with a uniquely shaped vocal instrument

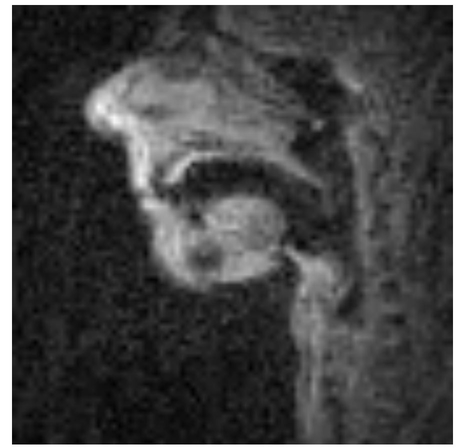


Novel biometric systems

S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd. An approach to real-time magnetic resonance imaging for speech production. J. Acoust. Soc. Am., 115:1771-1776, 2004.

Clinical Applications..





Cancer/Oral tongue
Post-Glossectomy Scans

Ongoing:

Building one of the first low field dynamic MR imaging systems

End-to-End Innovation Pipeline:
from multimodal imaging to informatics through advances in signal processing and AI

Highlight 2

Computational Media Intelligence

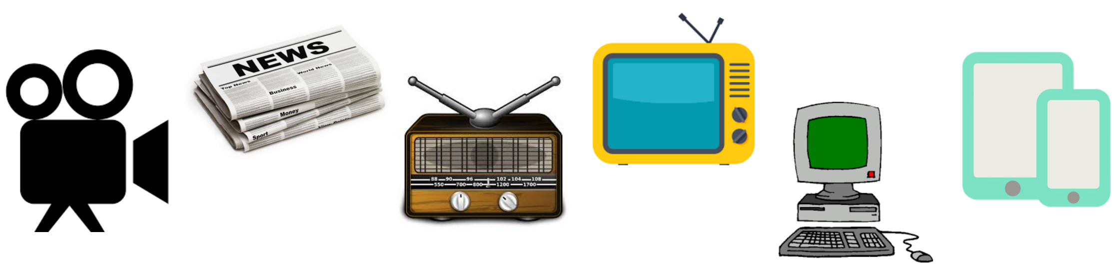
Special focus on Diversity and Inclusion

- 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

The Context: Creating & Experiencing Media Stories

- **An amazing range of domains**
  - to entertain, sell, educate, inform, influence,..

- **Across a variety of platforms**



9

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

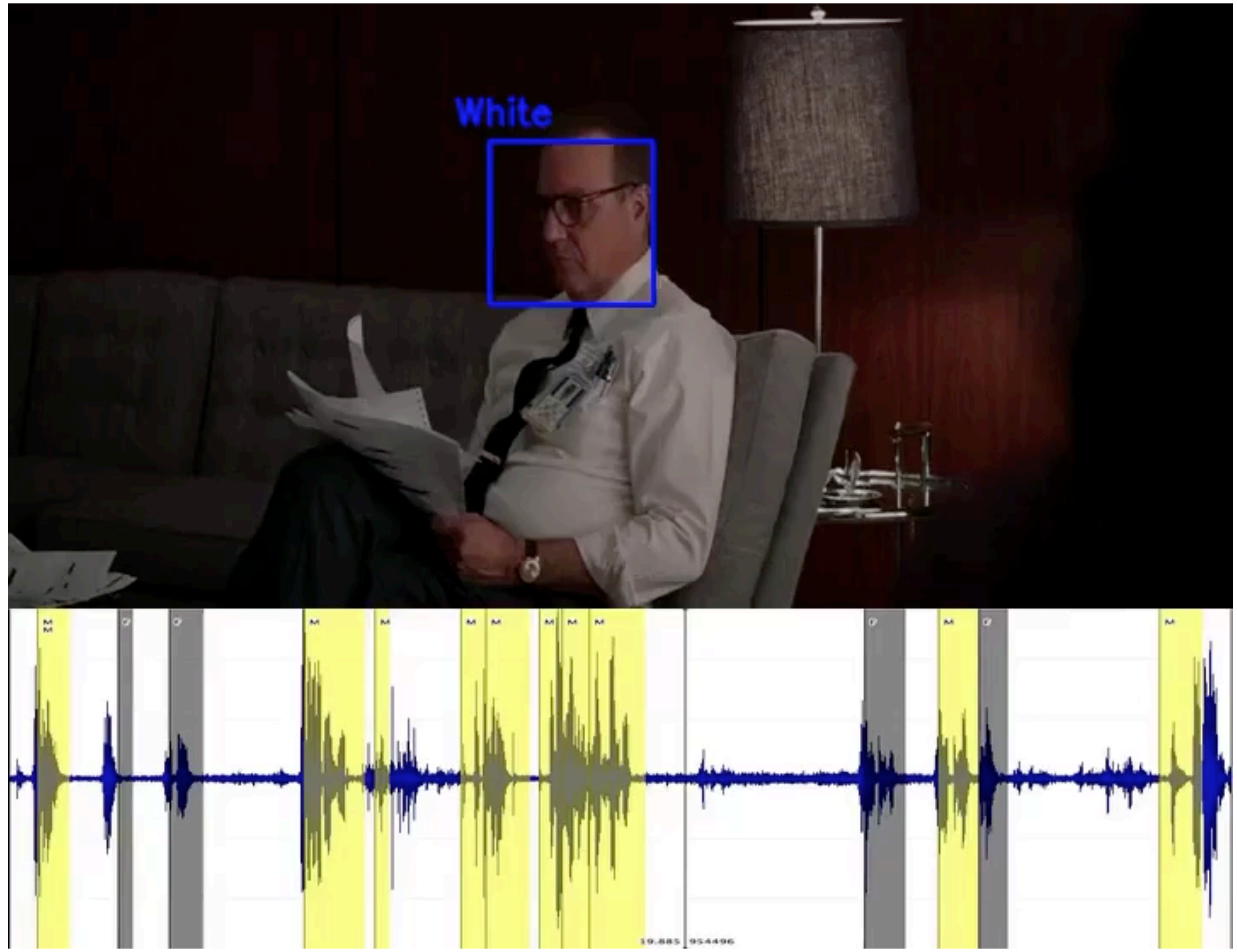
Screen-time

Female face

Male face

Race (from face)

(A) African-
(I) american
(W) Indian-Asian
(E) White
(L) East-Asian
( ) Latino/hispanic
unsure



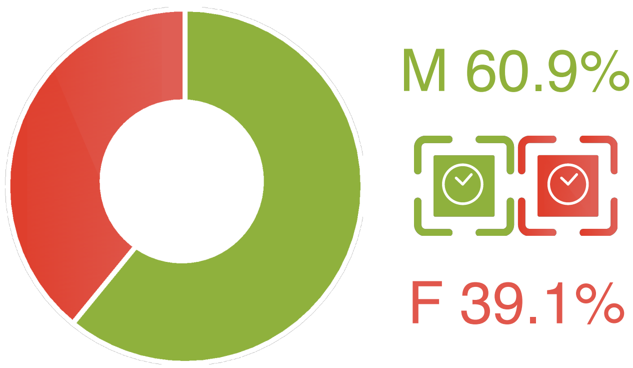
Speaking-time

Female speaker

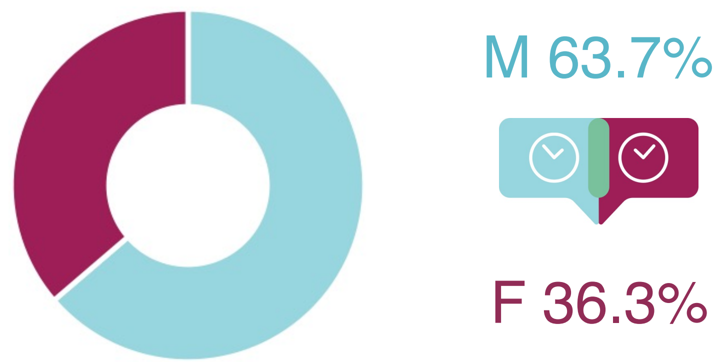
Male speaker

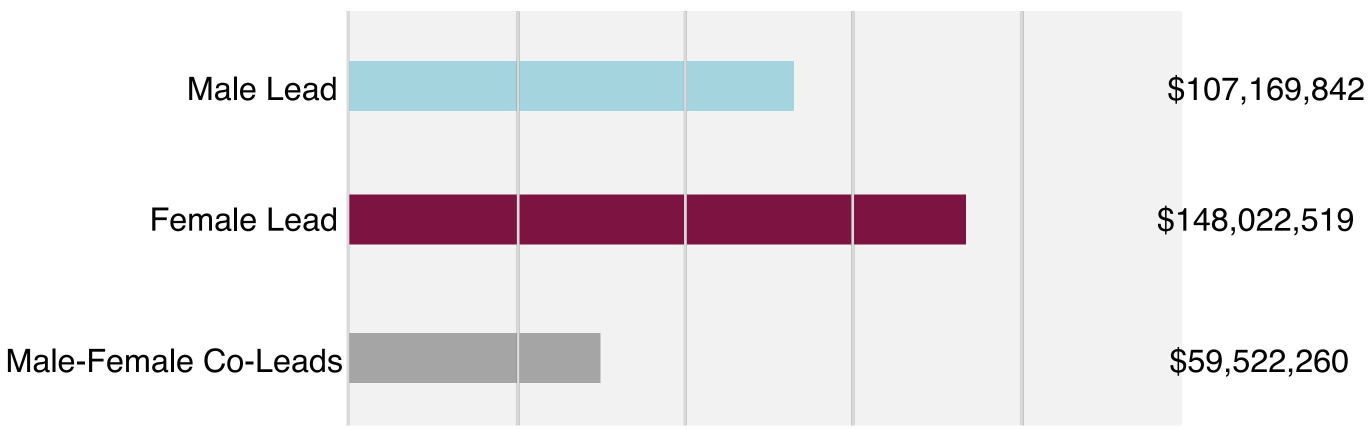
On top grossing ~100 live action US Films for 2017, 2018

2017 Screen Time by Gender (N = 94)



2017 Speaking Time by Gender (N = 94)





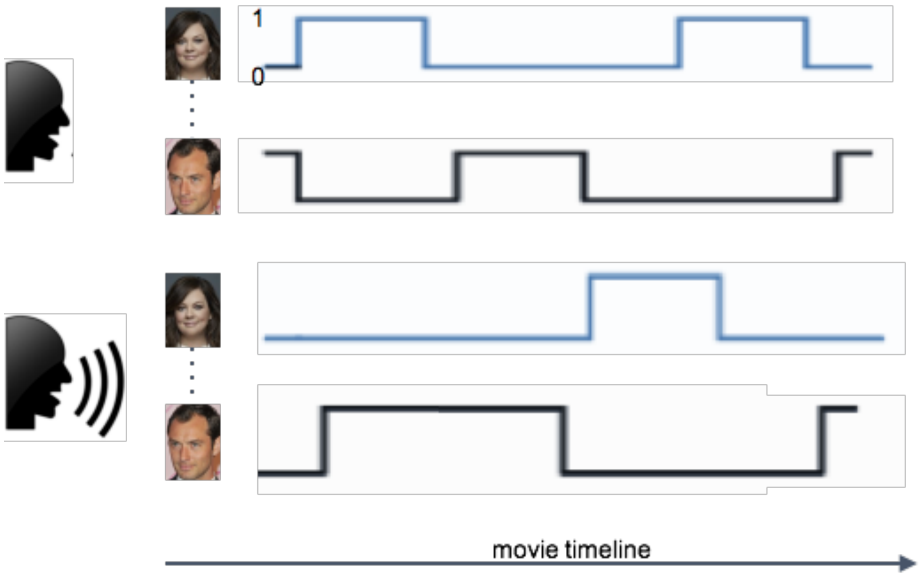
November 7, 2019: Analysis of 2.7 Million Ads—30% more views for ads with gender parity

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

Google

Joint Audio-visual Analysis: Sample insights

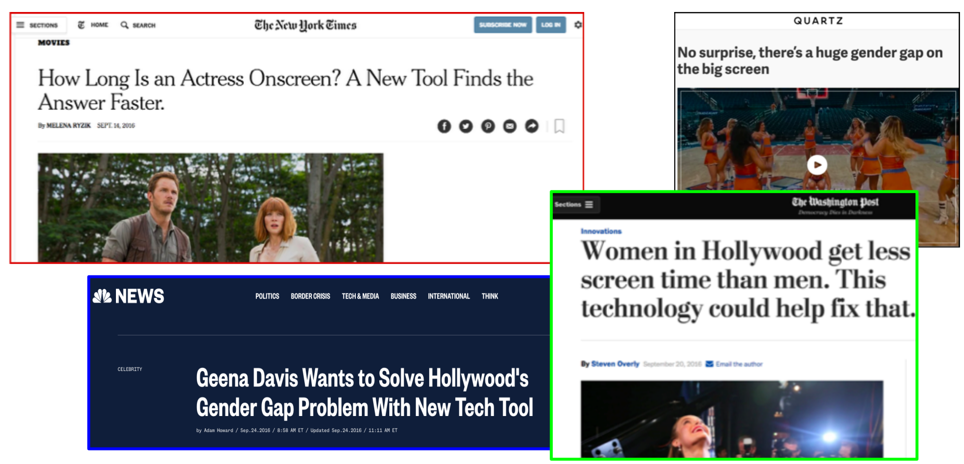
representational disparity



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*



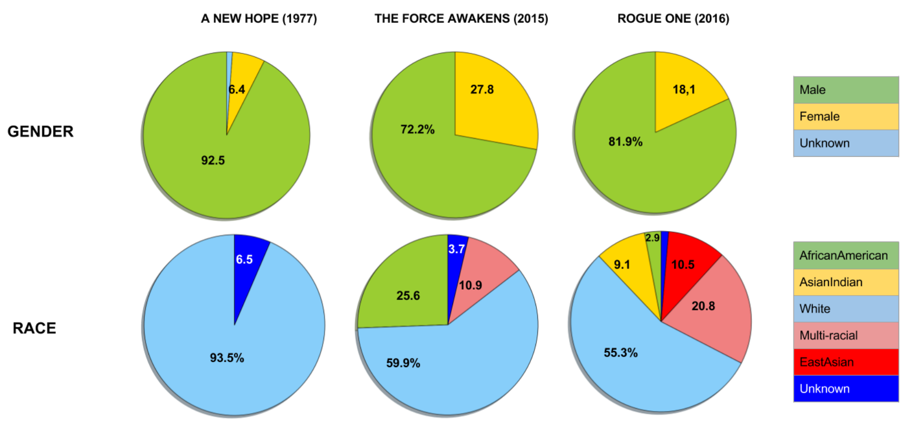
13

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

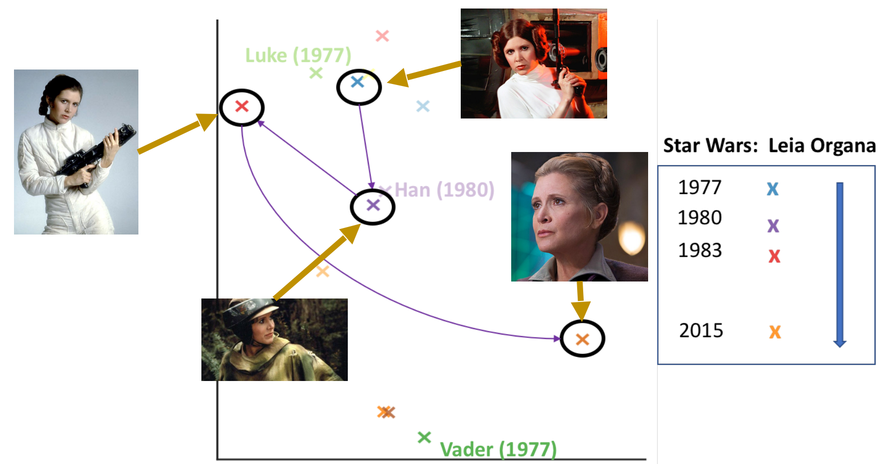
Text Analytics and Natural Language Processing

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




The New York Times

Look Who's Still Talking the Most in Movies: White Men

VOGUE

ARTS & ENTERTAINMENT

Female Characters In Films Often "Make No Difference To The Plot", Study Reveals

Los Angeles Times

MOVIES ENTERTAINMENT

USC study finds that movies are still dominated by men, on- and off-screen

THE TIMES

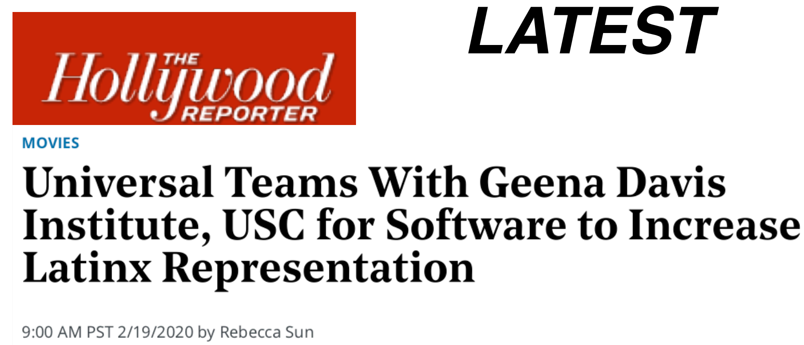
Why Hollywood's female stars get all the worst lines

INDEPENDENT

Culture

Female characters get all the worst lines in films, study says - despite making the most money in lead roles

LATEST


The Hollywood Reporter

MOVIES

Universal Teams With Geena Davis Institute, USC for Software to Increase Latinx Representation

9:00 AM PST 2/19/2020 by Rebecca Sun

14

Highlight 3

Behavioral Machine Intelligence and Mental Health and Well being

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 the human condition

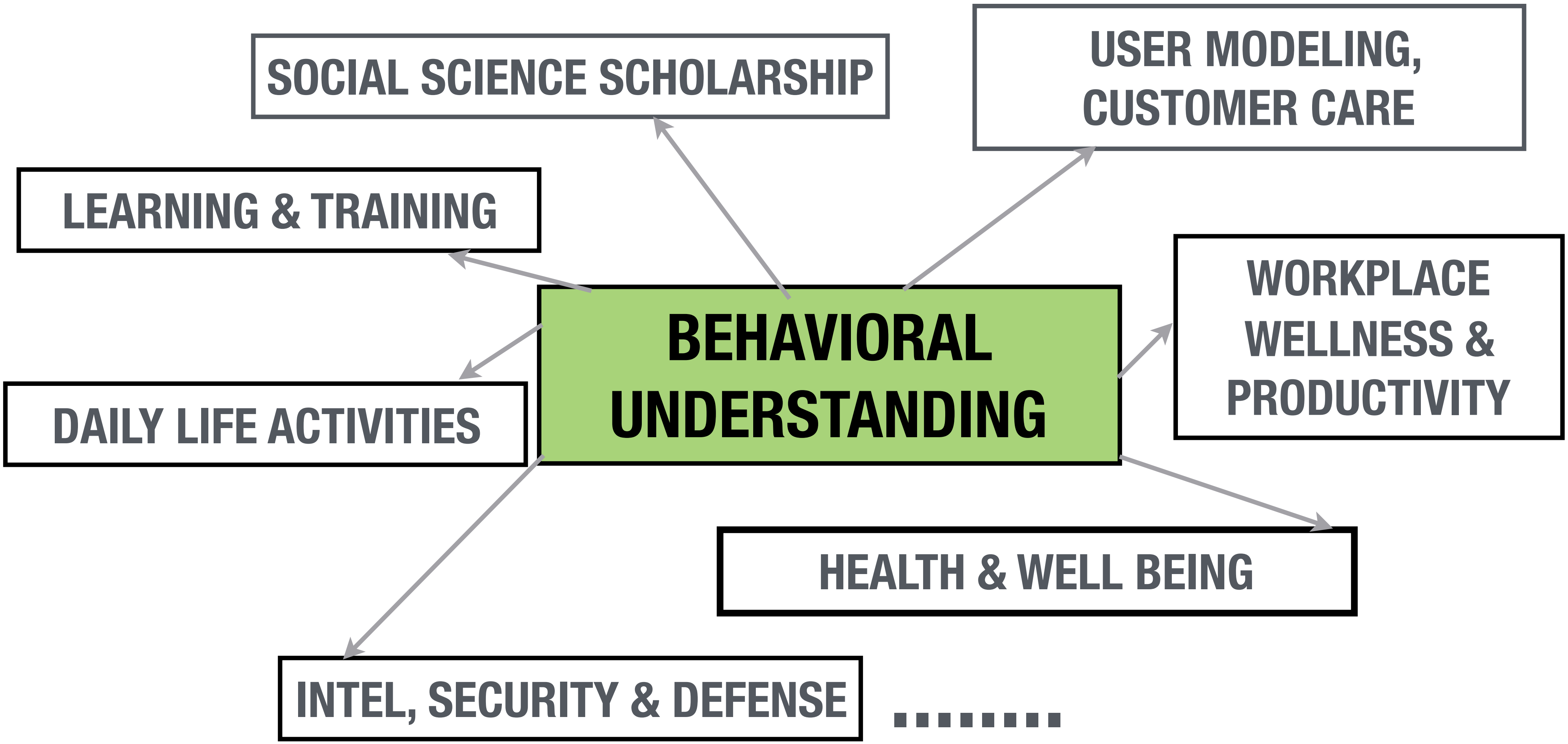
using engineering approaches and technologies

…..from qualitative to quantitative

Scalable, Broadly Accessible, Cost Effective

UNDERSTANDING BEHAVIOR CENTRAL TO MANY HUMAN ENDEAVORS

.... ACROSS APPLICATION DOMAINS



ROLE OF ENGINEERING?

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	> 80 years old	1.9% (lifetime)
Dementia (e.g., Alzheimer's disease)	> 60 years old	6.5% (lifetime)

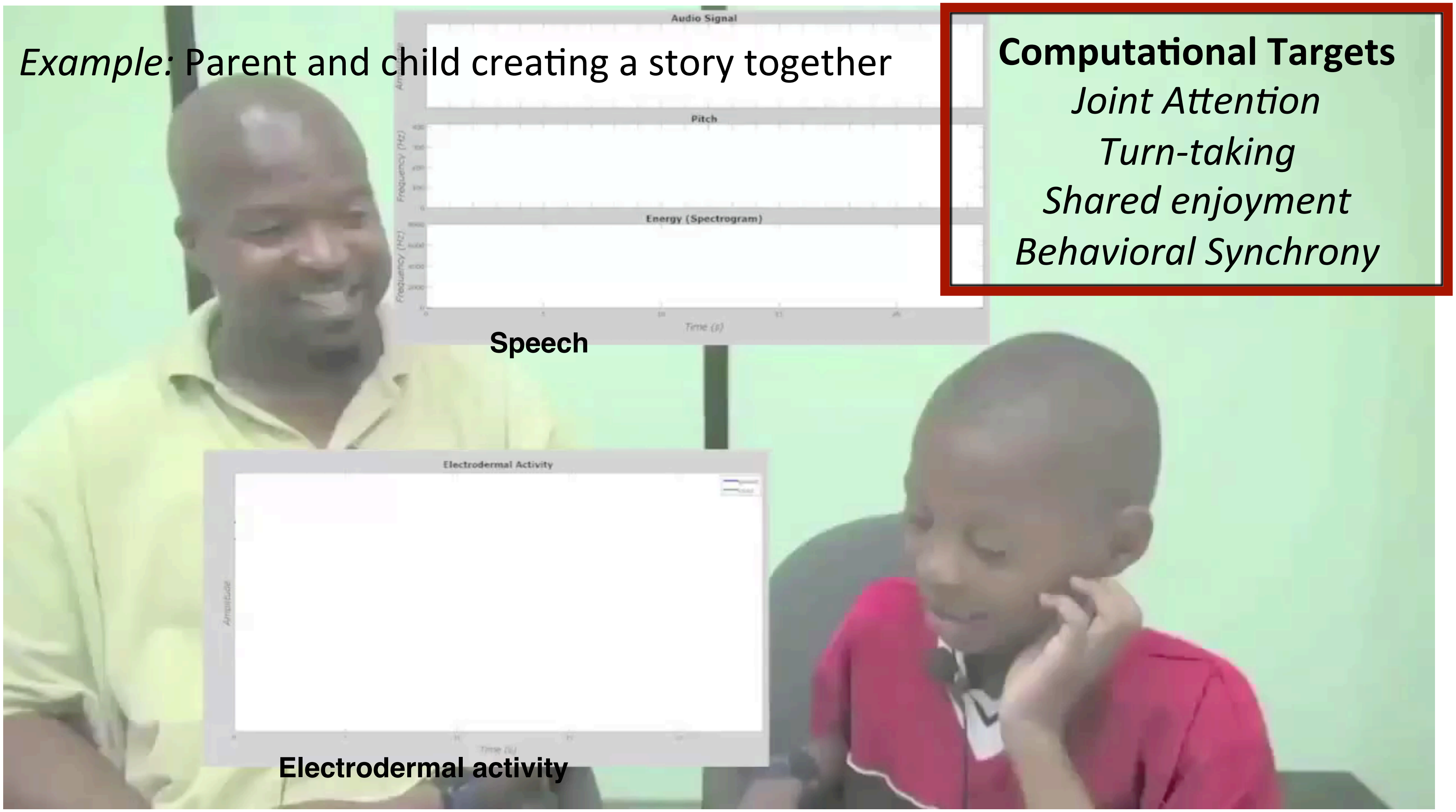
***Sources listed in:**

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

NIH, Simons Foundation

Autism Spectrum Disorder

Technologies for Rich Understanding of Expressive Behavior and Interaction

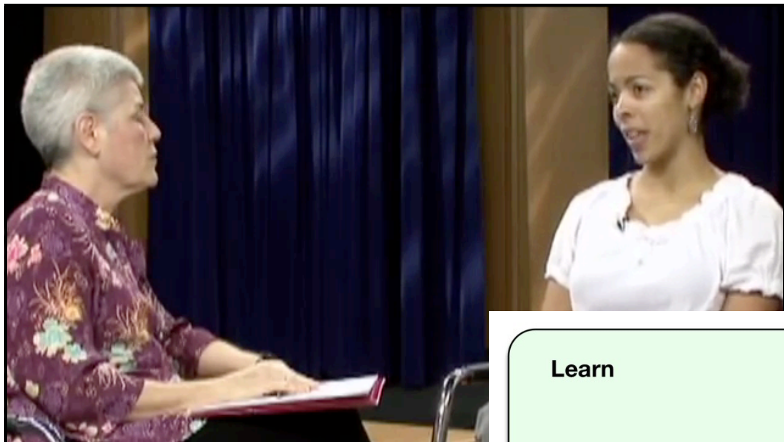


Economic Annual Cost of ASD in the US: $11.5 billion – $60.9 billion (2011 Dollars)

CDC https://www.cdc.gov/ncbddd/autism/data.html

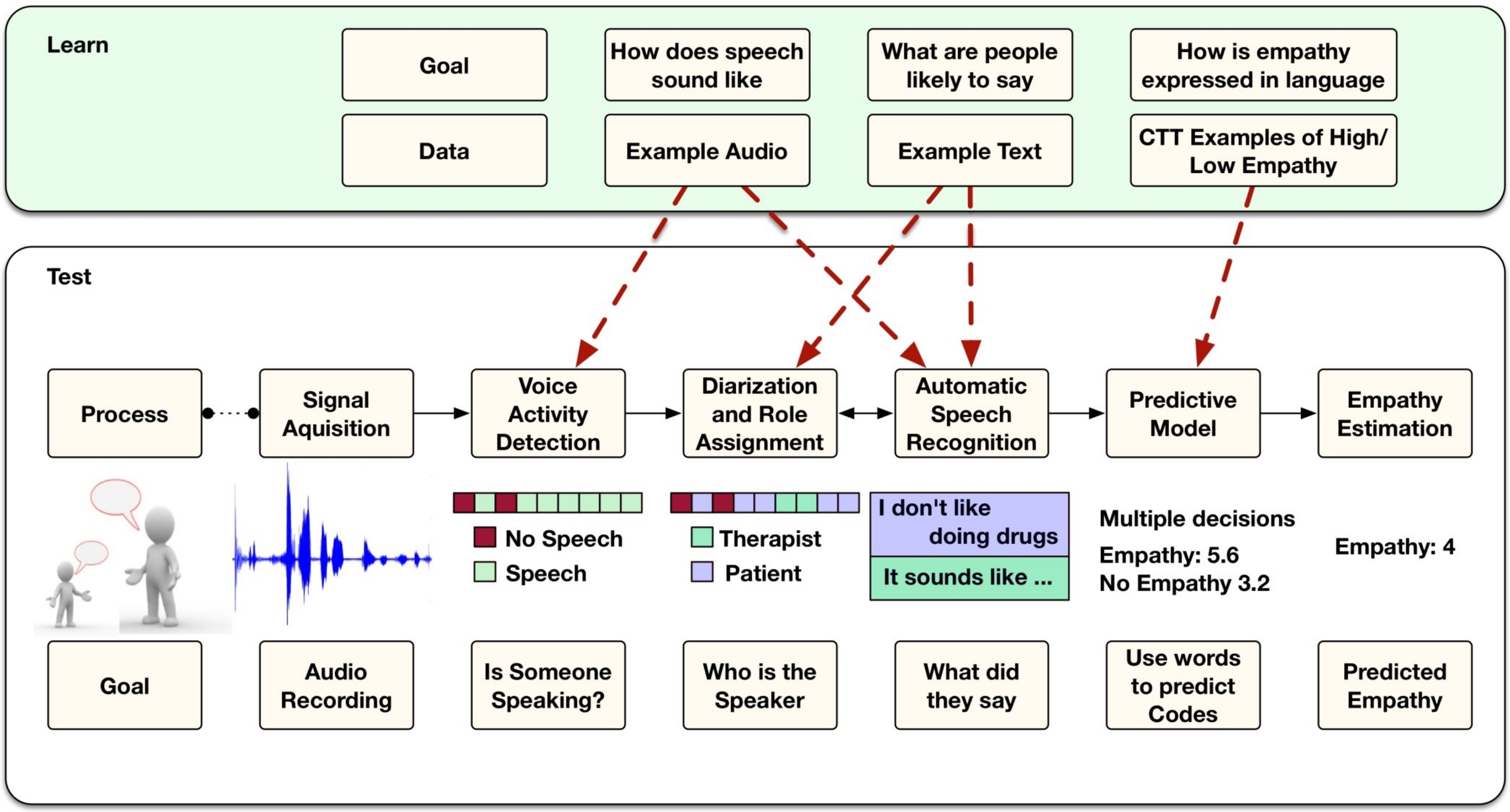
Addiction treatment: Psychotherapy

Illuminating what works, for whom, how and why



Motivational Interviewing

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



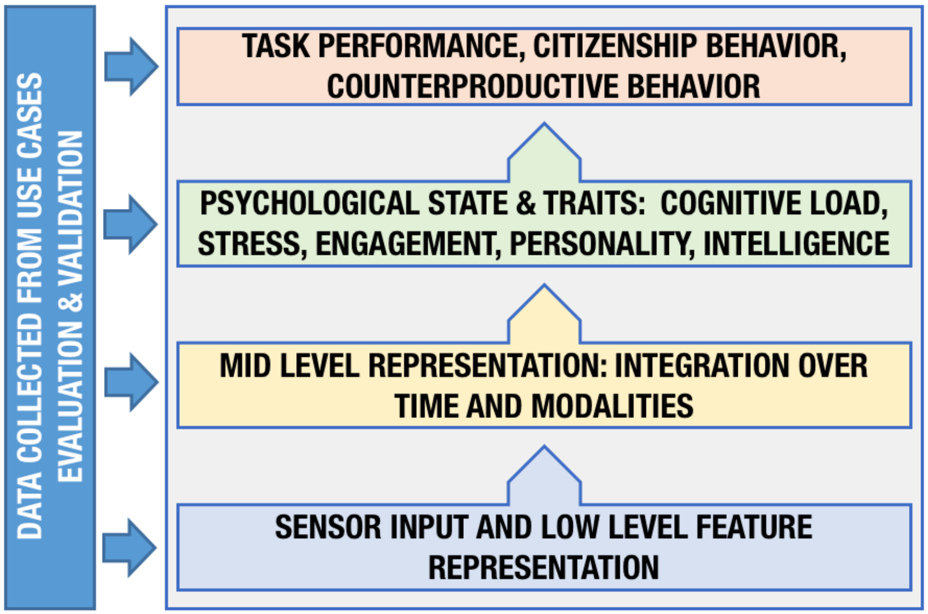
Annual costs of addiction exceed $740 Billion

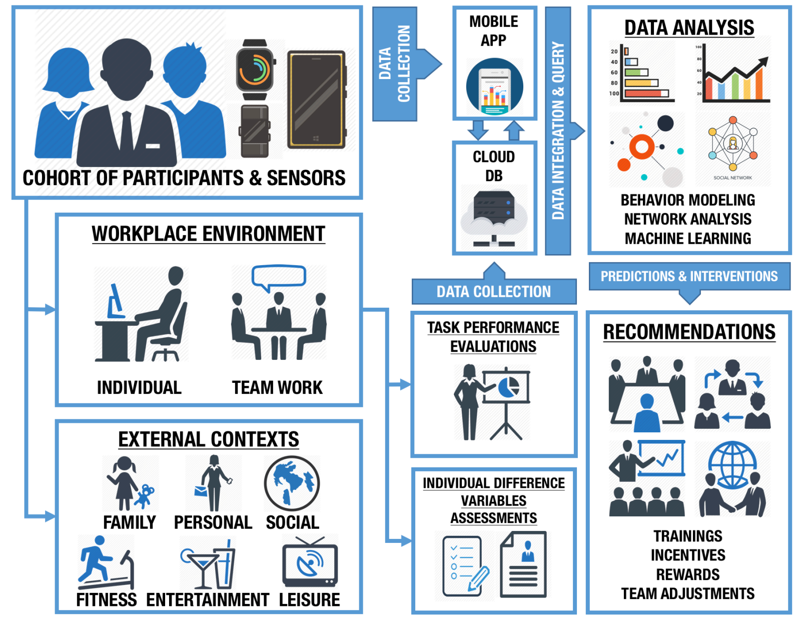
https://www.drugabuse.gov/related-topics/trends-statistics

IARPA, NSF

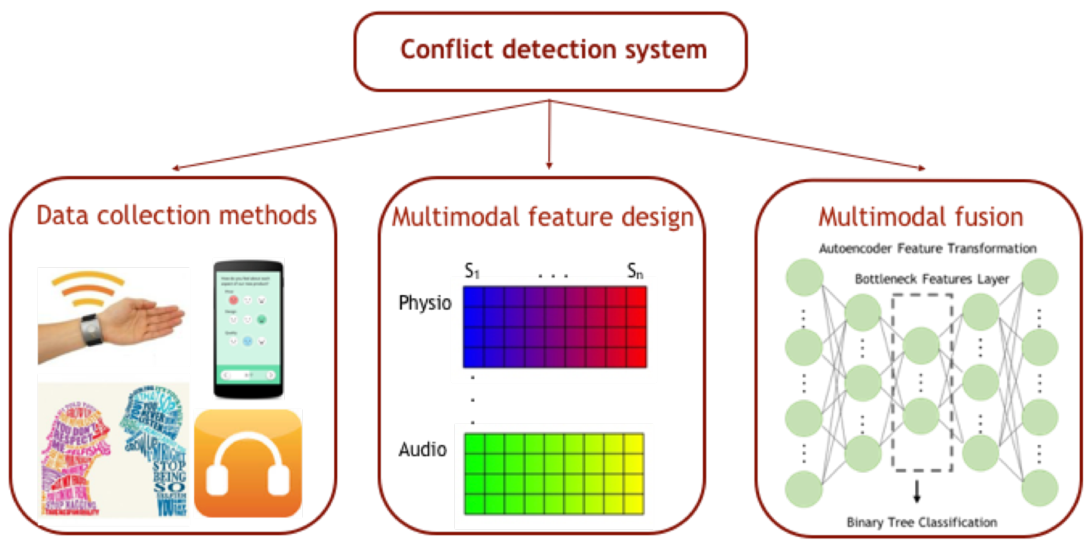
Health, Well being and Work place productivity

Bio-behavioral & IoT platform for individualized performance assessment





Conflict? Stress?



BEHAVIORAL 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

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*: e.g., location, temperature, light, sound, humidity, air qlty,..
- ***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, 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

How is technology helping already?

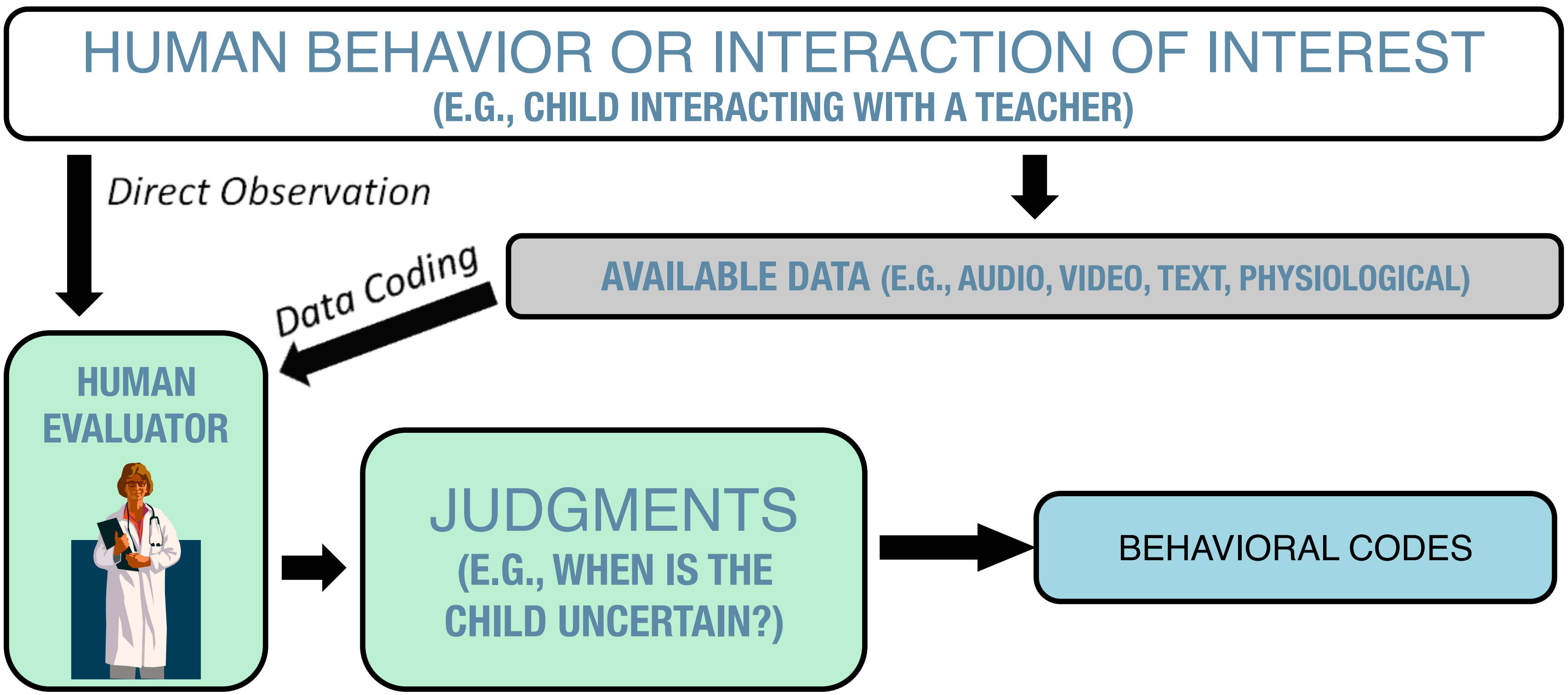
Sensing, Signal processing and machine learning are key enablers

- **Significant advances in wearable and context sensing**
- **Significant advances in foundational technologies for 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, ..
- **Significant advances in multimodal affective computing**

SHIFT TO MODELING MORE ABSTRACT, DOMAIN-RELEVANT HUMAN BEHAVIORS
......NEEDS NEW MULTIMODAL & MODELING APPROACHES

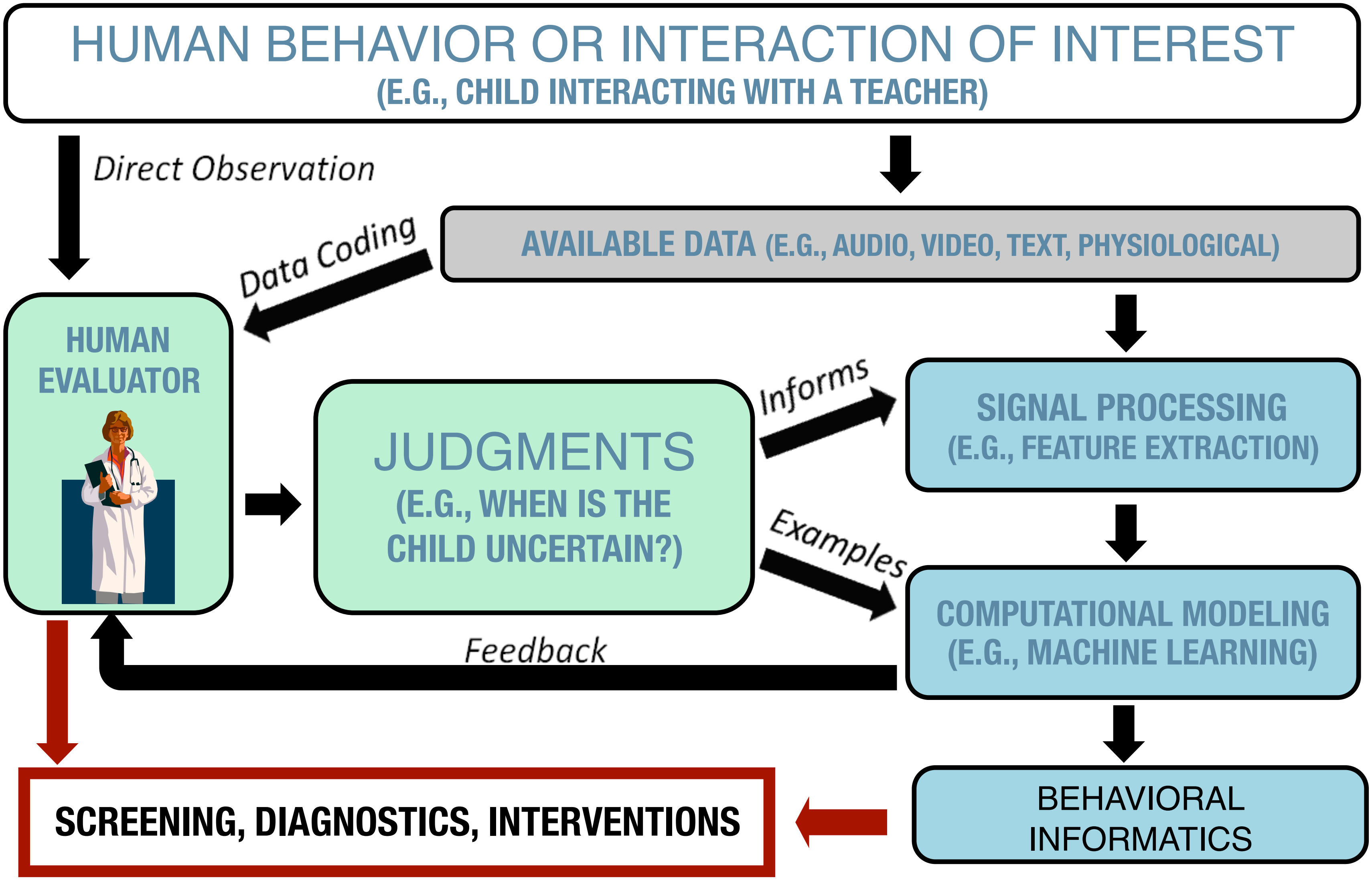
Behavior Coding: Humans in the loop

- Human assessments/judgments on human behavior



Behavior Coding: Humans in/on the loop

- Support—than supplant—human (expert) analyses



Collaborative integration human and machine intelligence

26

Behavioral Machine Intelligence: Human centered

COMPUTING

OF	human action and behavior 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

Two Illustrative Case Studies

- Autism Spectrum Disorder ***Diagnostics***
  - Characterizing and quantifying behavioral phenotypes
  - Technologies for personalized interventions
  - Machine learning for Dx
- Addiction ***Intervention***
  - Understanding and evaluating psychotherapy

Autism Spectrum Disorder (ASD)

- **1 in 54 US children diagnosed with ASD (CDC, 2016)**
  - 1% prevalence in Asia, Europe, North America, 2.6% in S. Korea
- **ASD characterized by**
  - Difficulties in social communication, reciprocity
  - Repetitive or stereotyped behaviors and interests

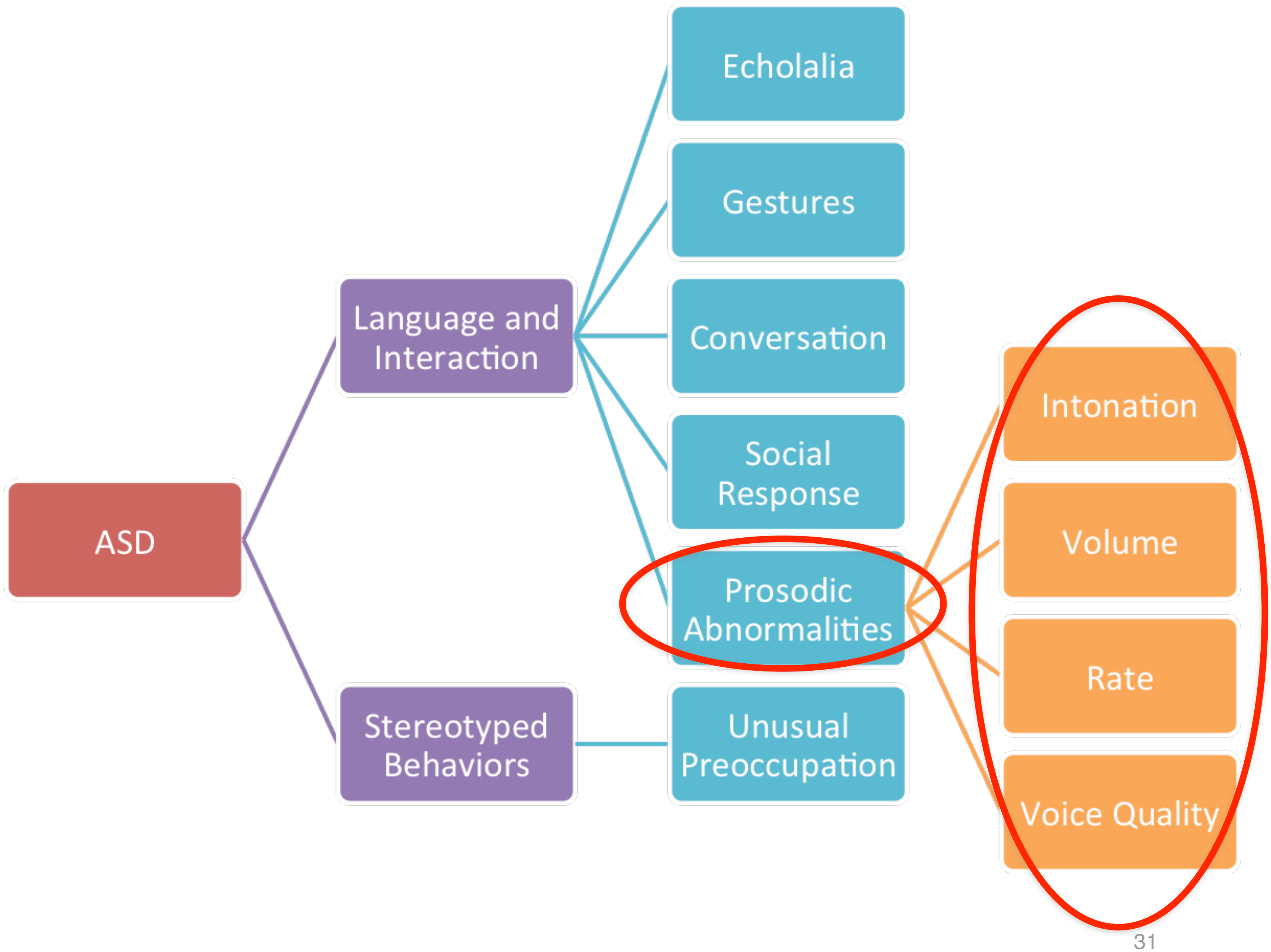
Technology possibilities include

- Illuminating social communication and behavioral patterns
- Stratifying phenotypes with objective and adaptable metrics
- Track, quantify behavior change (e.g., response to interventions)
- Technologies to support measurements and intervention delivery: *personalized, just in time, ecologically valid*

Analyzing Interaction in ASD

- **Assessment, Intervention, Game play/training Examples**

ASD Assessment



Quantifying Atypical Prosody

Qualitative descriptions are general and contrasting

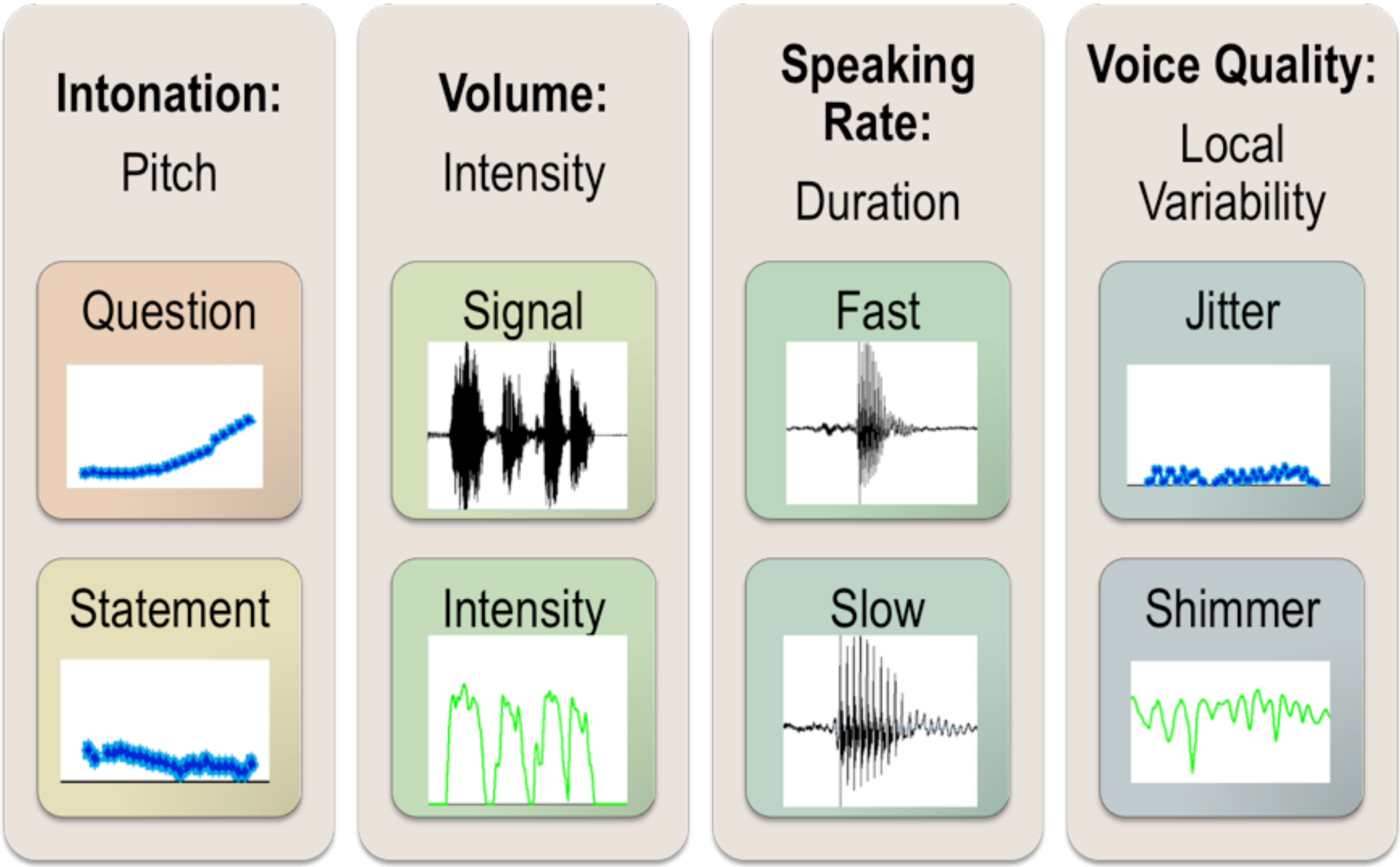
ADOS

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

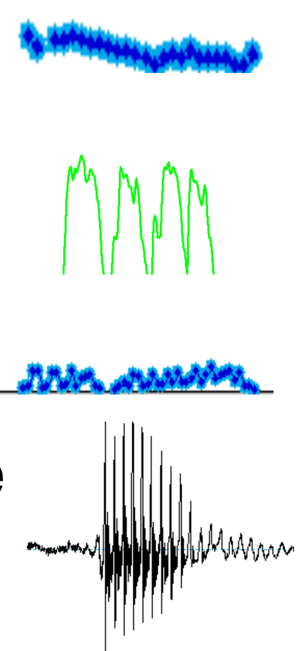
✦ ***median, IQR of above***

Atypical Prosody & Interaction

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

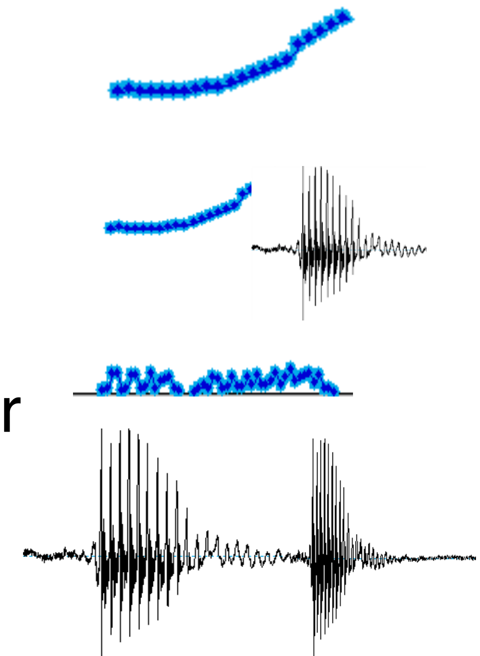
Child's Prosody

- "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, Matthew P. Black, Chi-Chun Lee, Marian E. Williams, Pat Levitt, Sungbok Lee, and Shrikanth Narayanan, "Spontaneous-Speech Acoustic-Prosodic Features of Children with Autism and the Interacting Psychologist", Interspeech, 2012.

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.

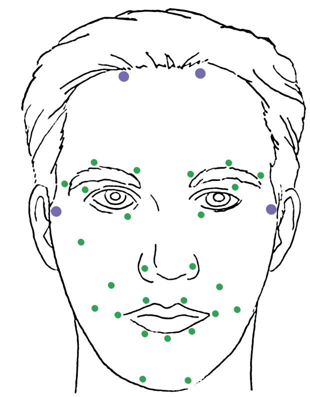
ASD: Understanding the expression of social cues

Example:
Production of Affective Facial Expressions (During Smile Imitation Task)

Reduced complexity in dynamic facial behavior primarily in the eye region

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

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



Objective

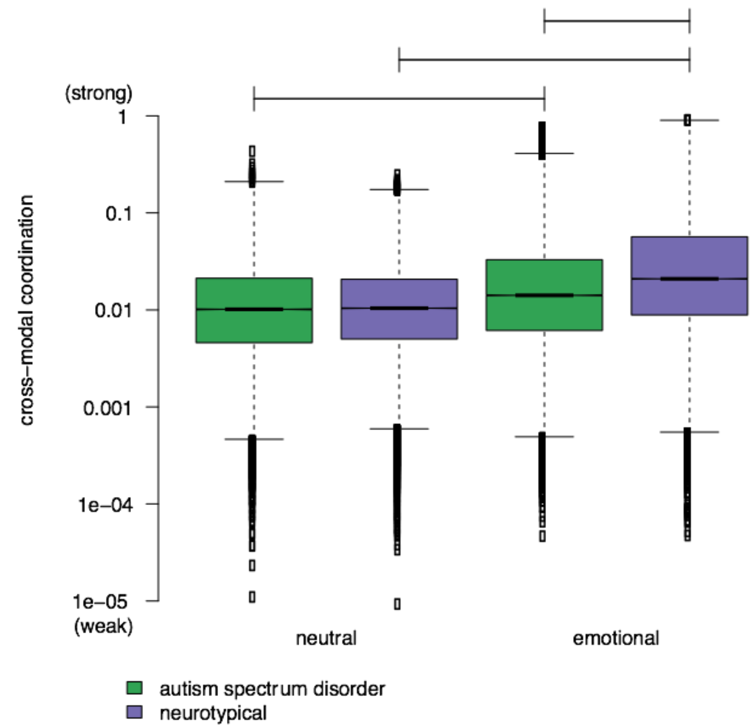
- Dynamic relation between speech production and facial expression in children with autism?
- How face-directed gaze modulates this cross-modal coordination?

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

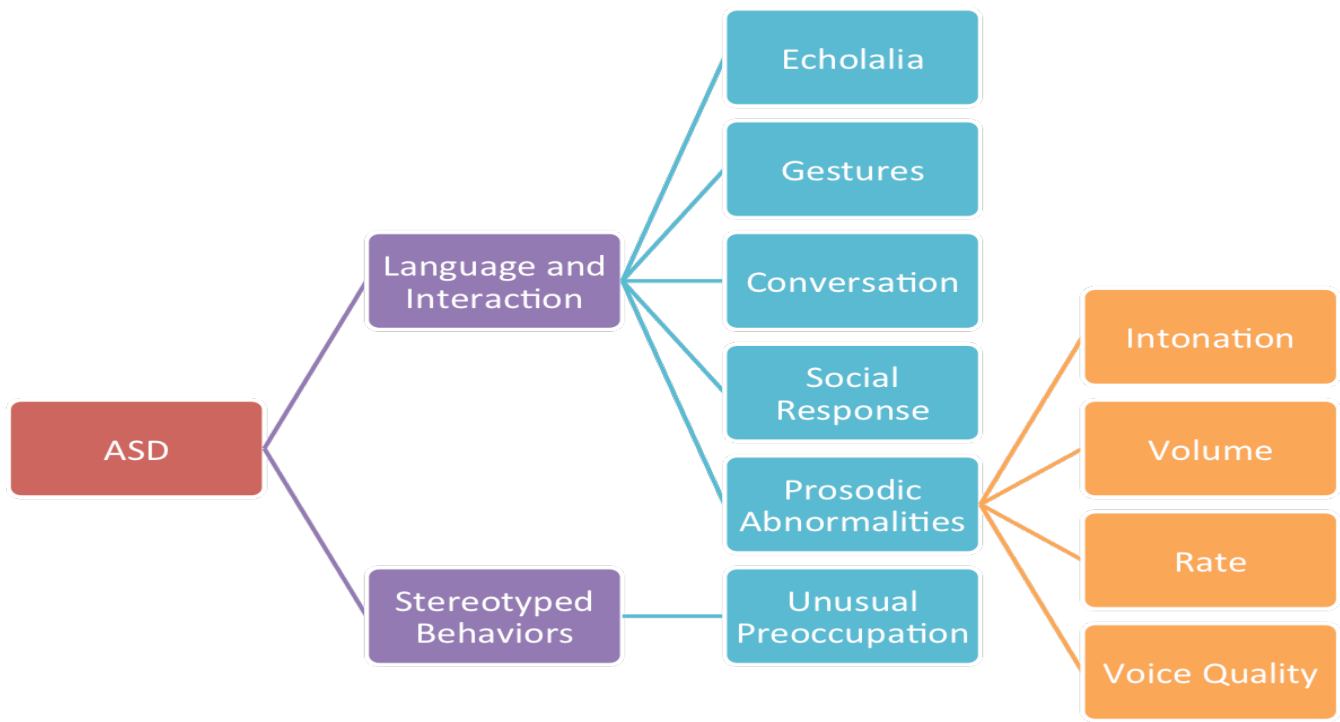
- Neurotypical children produced emotional sentences with strong cross-modal coordination and produced neutral sentences with weak cross-modal coordination
- Autistic children produced similar levels of cross-modal coordination for both neutral and emotional sentences.
- Cross-modal coordination was greater when the neurotypical 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

Opportunities for rich multimodal approaches

- 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

Some Case Studies

- Dyadic interaction and relationship dynamics ***Modeling***
  - Behavioral Coding
  - Modeling Interaction dynamics
  - Conflict
- Autism Spectrum Disorders ***Diagnostics***
  - Characterizing and quantifying behavioral phenotypes
  - Technologies for personalized interventions
  - Machine learning for Dx

✓ Addiction ***Interventions***

  - Understanding and evaluating psychotherapy

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

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

Human observation & annotation

→ Therapist Empathy ratings

Automatic Inference

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

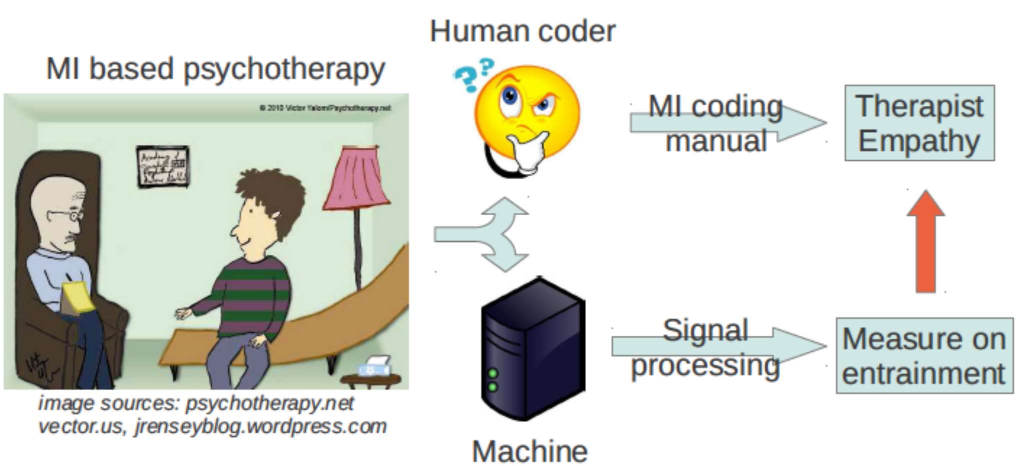
Key Findings

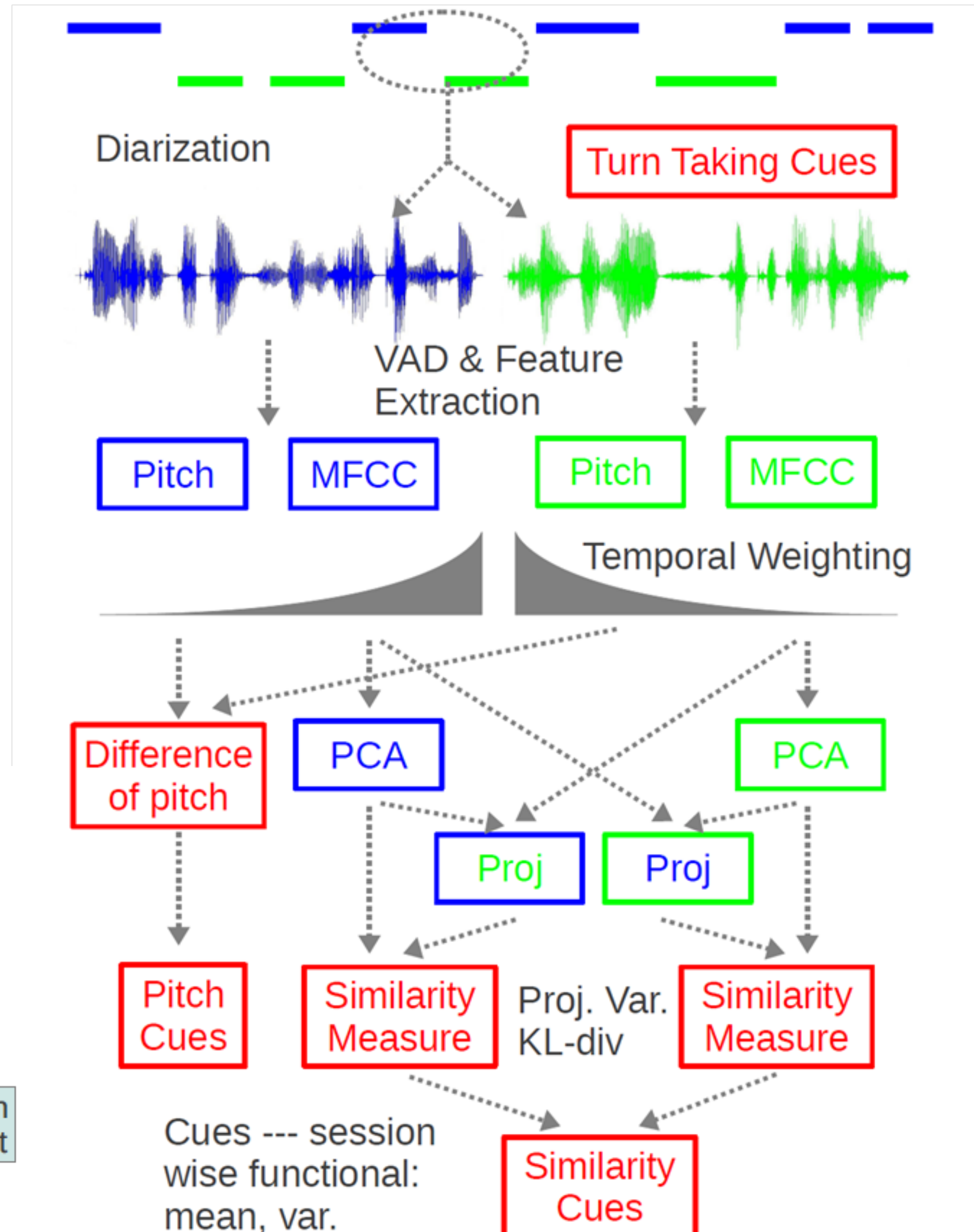
- **lower perceived empathy of therapist when:**
  - Therapist has higher energy values
  - Therapist has higher pitch values

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**



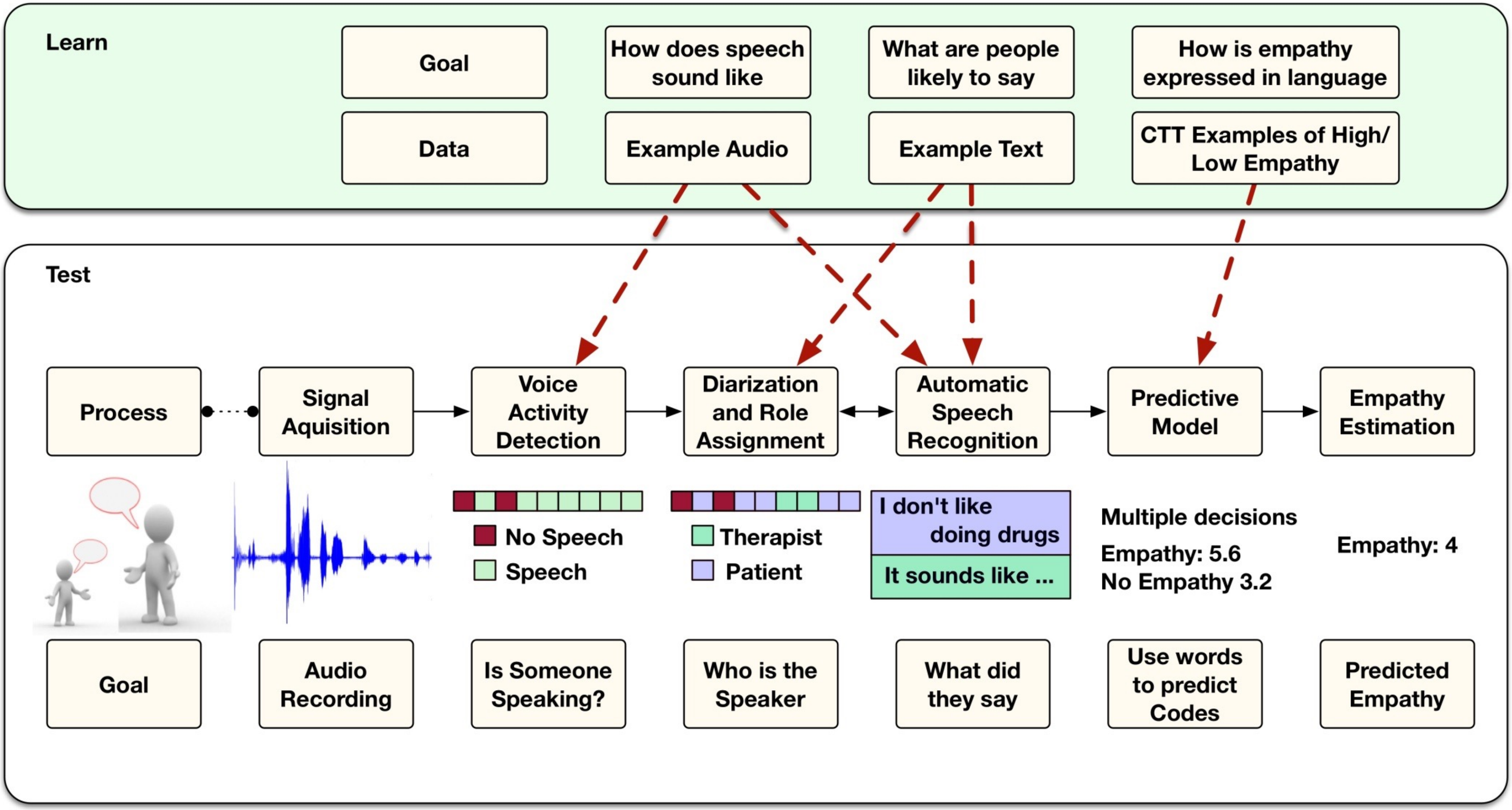


Bo Xiao, et al., Modeling Therapist Empathy and Vocal Entrainment in Drug Addiction Counseling Proceedings of Interspeech, 2013

“Sound to code” system:

Used by 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

Open Challenges

- **Getting the right multimodal data**
  - *sensing* in natural context; 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, integrity, sharing, and management
  - Enabling interventions & evaluation at scale, cost, JIT
  - Choosing the right operating point: adaptivity

Concluding Remarks:
Enabling Behavioral informatics

- **Human behavior can described from a variety of perspectives**
  - Both challenges ***and*** opportunities for R&D
  - Multimodal data integral to derive and model these features
- **Computational advances: sensing, processing and modeling**
  - Support **BOTH** human and machine decision making
- **Exciting technological and societal possibilities**
  - Opportunities for interdisciplinary and collaborative scholarship

BEHAVIORAL MACHINE INTELLIGENCE AND INFORMATICS:
COMPUTING BEHAVIORAL TRAITS & STATES FOR DECISION MAKING AND ACTION

✓Helps do things we know to do well more efficiently, consistently
✓Helps handle new data, create new models to offer unimagined insights
✓Creates tools for discovery

Work reported represents efforts of <u>numerous</u> colleagues and collaborators

Too many to name, but grateful to all

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

SHRIKANTH NARAYANAN AND PANAYIOTIS GEORGIOU. BEHAVIORAL SIGNAL PROCESSING: DERIVING HUMAN BEHAVIORAL INFORMATICS FROM SPEECH AND LANGUAGE. PROCEEDINGS OF IEEE. 101(5): 1203 - 1233, 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

USC School of Engineering

http://sail.usc.edu/

University of Southern California