Multimodal Behavioral Informatics

Creating technologies to work with and enhance human capabilities and experiences

Shrikanth (Shri) Narayanan

Signal Analysis and Interpretation Laboratory (SAIL)

http://sail.usc.edu

University of Southern California

CREST SYMPOSIUM, APRIL 2012
KYOTO UNIVERSITY

1

Human Behavior

Complex and multifaceted: involves

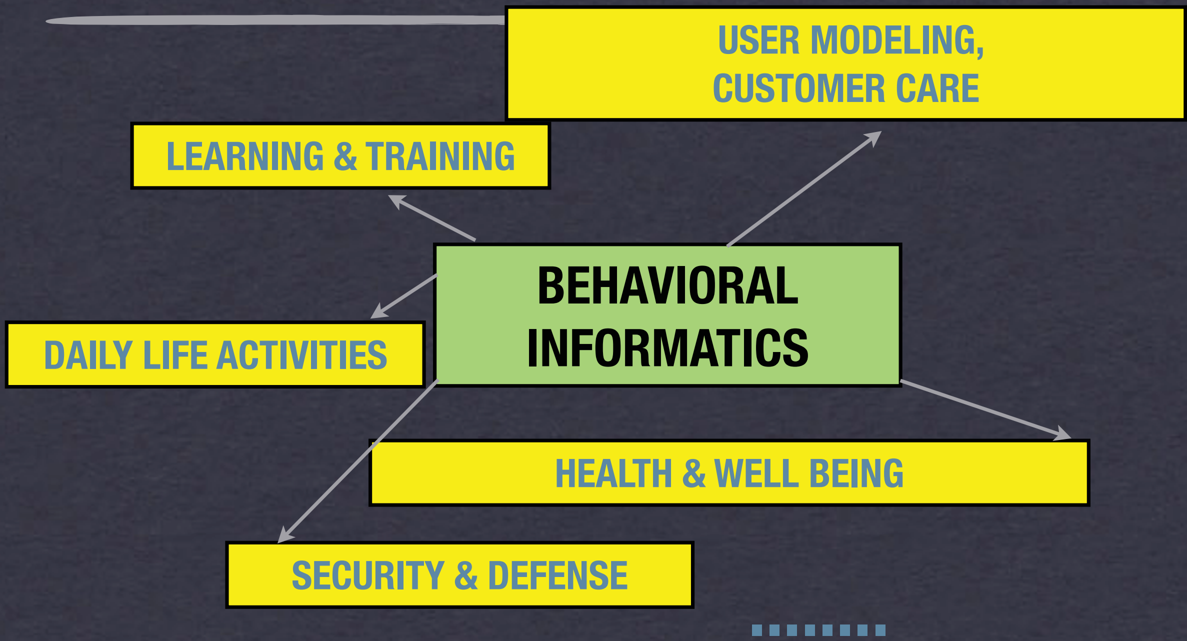
- Intricate brain-body interplay
- Effect of environment and interaction with others

- Communication, Affect, Personality, Social interaction,..
- Generation and processing of multimodal cues
- Typical, Atypical, and Disordered characterizations

ROLE OF ENGINEERING?

2

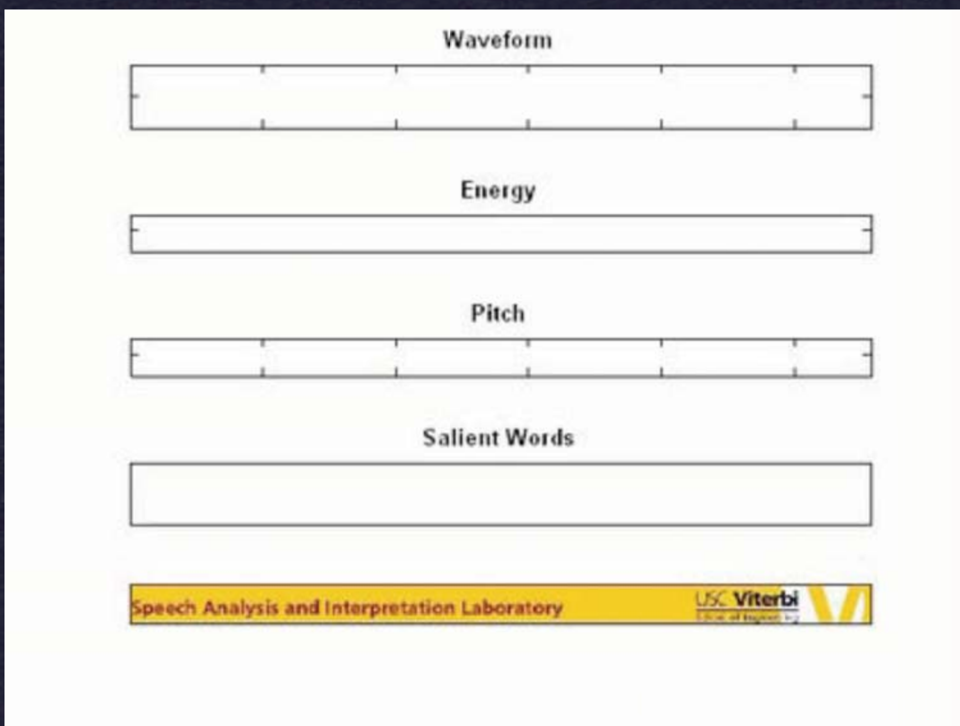
ACROSS NUMEROUS DOMAINS BEHAVIOR ANALYSIS IS CENTRAL



BUT LARGELY HUMAN BASED!

3

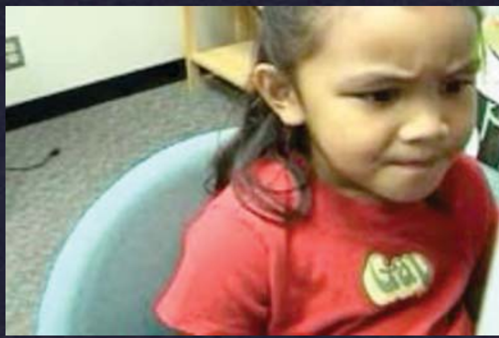
Customer care
Is the customer on the telephone upset? (only customer side played)



4

Educational Game: “Cognitive state” Characterization

- **CONFIDENT** vs. **UNCERTAIN**



“Uncertainty” manifests itself through combination of vocal, language, and visual behavioral cues

5

Marital Couple therapy
Characterizing affective dynamics, blame patterns

CHRISTENSEN/CIRS TRAINING

6

Autism Spectrum Disorders
Characterizing joint attention; shared enjoyment

Virtual Human Interactions
Training behavior: doctor-virtual patient interactions

Multimodal Behavior Signals

✓ Provide a window into higher level processes

- Some overt and directly observable
  - e.g., vocal and facial expressions, body posture
- Others, covert
  - e.g., heart rate, electrodermal response, brain activity

✓ Implications for

- ✓ Understanding "mind-body" relations
- ✓ Understanding people's judgment of others behavior

MEASURING & QUANTIFYING HUMAN BEHAVIOR:
A CHALLENGING ENGINEERING PROBLEM

9

Operationally defining

Behavioral Signal Processing (BSP)

COMPUTATIONAL METHODS THAT MODEL HUMAN BEHAVIOR SIGNALS

- MANIFESTED IN BOTH OVERT AND COVERT CUES
- PROCESSED AND USED BY HUMANS EXPLICITLY OR IMPLICITLY
- FACILITATE HUMAN ANALYSIS AND DECISION MAKING

- OUTCOME OF BSP: "BEHAVIORAL INFORMATICS"

QUANTIFYING
HUMAN EXPRESSED BEHAVIOR
AND
HUMAN "FELT SENSE"

10

How technology has helped already?

- **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 Affective signal processing with EKG, GSR, ..

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

Ongoing Advances: Multifaceted

- **Sensing: From Smartrooms to Body area networks**
- **Rich speech/spoken language and video understanding**
  - ***who* said *what* to *whom, how* and *when & where***
- **Affective computing & Emotion recognition**
  - Modeling affective behavior in acted and natural scenarios
- **Social signal processing**
  - Modeling individual and group social behavior: turn taking, non verbal cues such as smiles, laughters and sighs, head nods, proxemics, ...

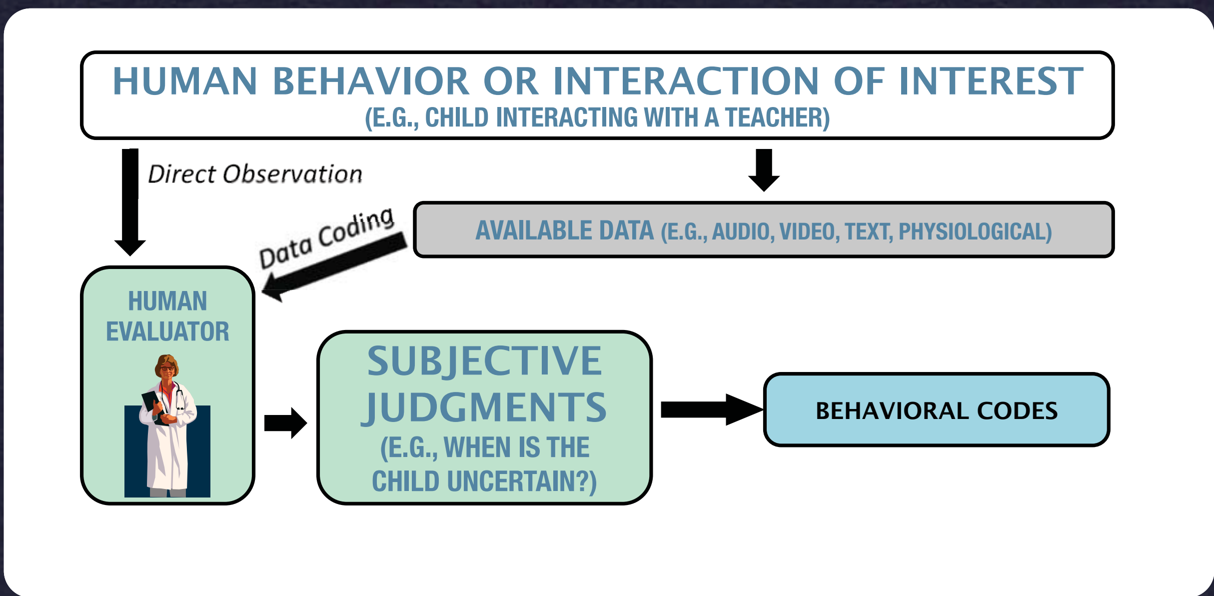
ALL THESE ARE ESSENTIAL BSP BUILDING BLOCKS:
“LOW & MID LEVEL DESCRIPTORS”

Behavioral Signal Processing: Ingredients

- **Acquisition: rich and ecologically valid data**
  - Behavior data sensing: audio, video, physiological, location,..
  - Measurements in controlled and natural free-living environments
- **Analysis: deriving signal descriptors**
  - Deriving low level cues: who, what, when, how, where, why
- **Modeling: mapping behavioral constructs**
  - High level descriptions desired by domain experts
  - Descriptive and predictive models using multimodal data
- **Handle varying types of abstraction in data and descriptions**
  - Uncertainty in observations (partial, noisy)
  - Subjectivity in descriptions (especially of higher level behavior)
  - Heterogeneity and variability in how data are generated and used

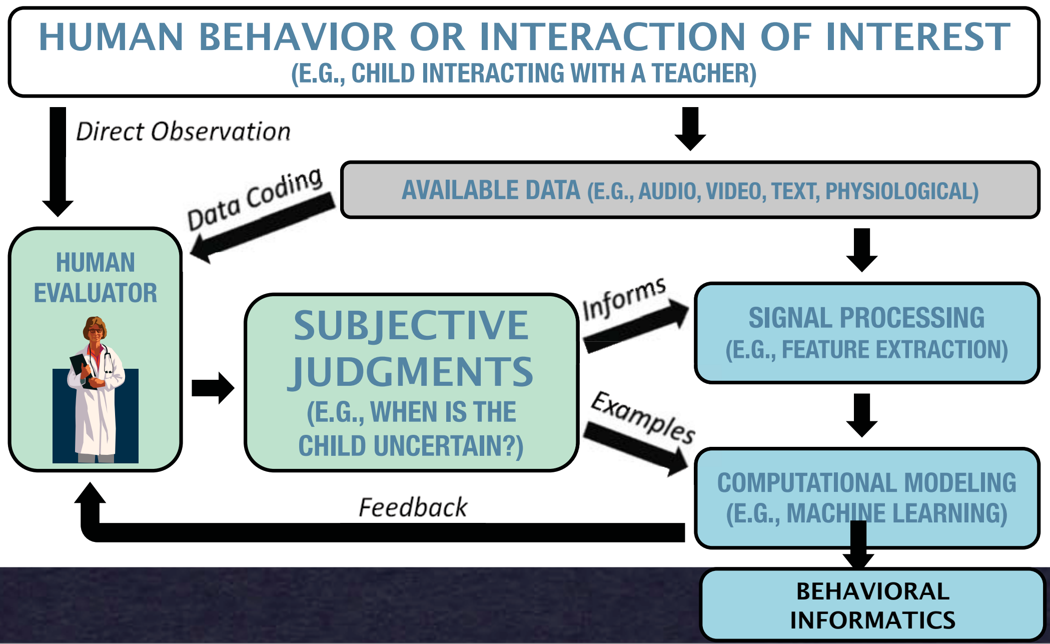
Behavior Coding: Humans in the loop

- Modeling subjective judgments on human behavior



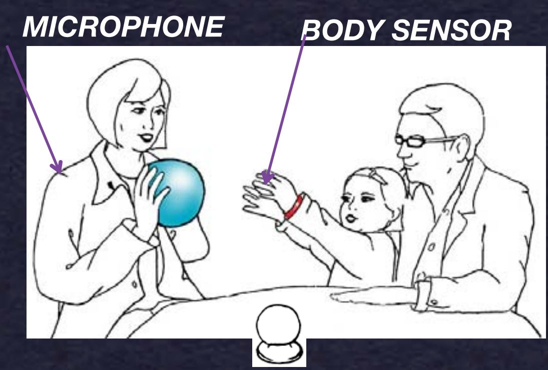
Behavior Coding: Humans in the loop

- Modeling subjective judgments on human behavior



15

Behavioral Signal Processing Pipeline: Summary



CHILDREN IN SCHOOL SETTING

DATA STREAMS:
AUDIO
VIDEO
PHYSIOLOGICAL

CAPTURING BEHAVIORAL SIGNALS
- FACE AND GAZE
- VOCALIZATION
- PHYSIOLOGICAL
- SYNCHRONIZATION
- ENVIRONMENT
- REFLECTION AND USABILITY

MEASURING BEHAVIORAL VARIABLES
- AFFECT
- ATTENTION
- ACTION AND SIGN RECOGNITION
- COMMUNICATIVE BEHAVIOR INVENTORY

UNDERSTANDING DYADIC BEHAVIORS
- BEHAVIOR CODING
- INTERACTION MODELING AND PARSING
- DIVERGENCE ANALYSIS
- VISUALIZATION AND RETRIEVAL

16

REST OF THIS TALK

- Some BSP building blocks

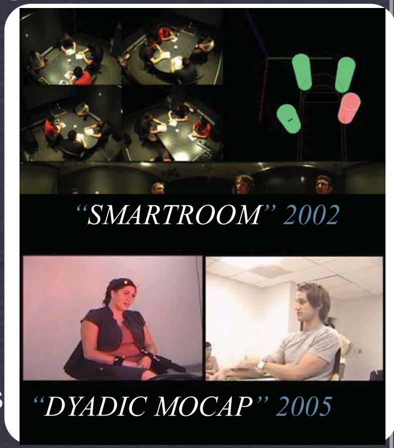
Example BSP Case Studies

- Family Studies: Marital couples
  - Blame patterns; positiveness/negativeness; humor/sarcasm
- Autism Spectrum Disorders
  - Characterizing and quantifying socio-emotional discourse
  - Technology interfaces for elicitation & personalized interventions
- Metabolic Health Monitoring
  - Technologies for ecologically valid measurements
  - Novel just in time analytics and intervention technology support

BSP BUILDING BLOCKS

Multimodal Signal Acquisition & Processing

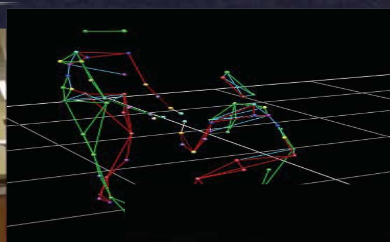
- **Instrumented Environments**: arrays of microphones, cameras, mocap
  - Sense the user
    - Identity/Location of speaker
    - Speech, visual activity
  - interaction
    - Turn taking
    - Back channels, head orientation, proxemics
  - and the audio-visual environment:
    - Lab, classroom, clinic, home, playground,..



- **Instrumented People**: Body sensing, mobile settings
  - Sense user state, activity, context
  - Wireless (cell phone) based: sensing & actuation,
  - Data from real life free living conditions

19

EXPRESSIVE BEHAVIOR: USING ACTORS



USC CreativeIT database

- **Multimodal** database
- USC Engineering and Theatre
- Dyadic Theatrical Improvisations
- Motion Capture, Video, Audio
- http://sail.usc.edu/improv/

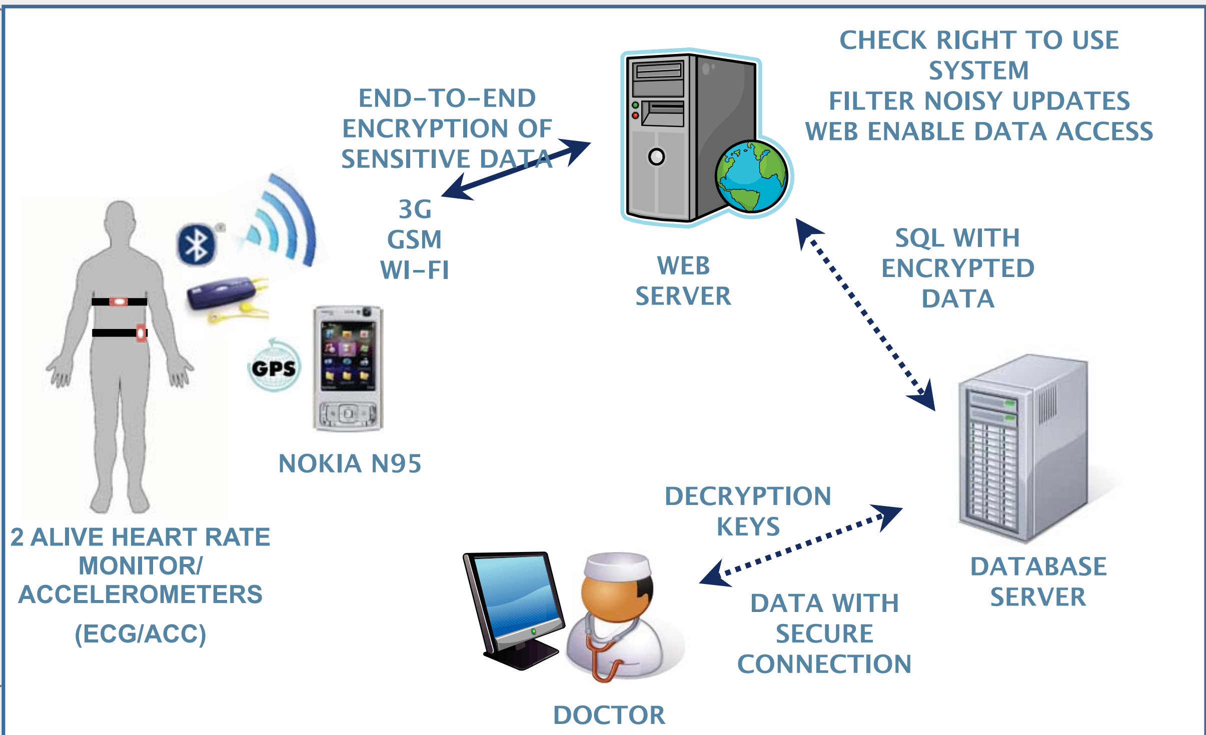
USC IEMOCAP: Interactive and emotional motion capture database

- Dyadic interaction
- 5 sessions, 2 actors each
- Emotions elicited in context
- ~12 hours of data
- http://sail.usc.edu/iemocap/

FREELY AVAILABLE

20

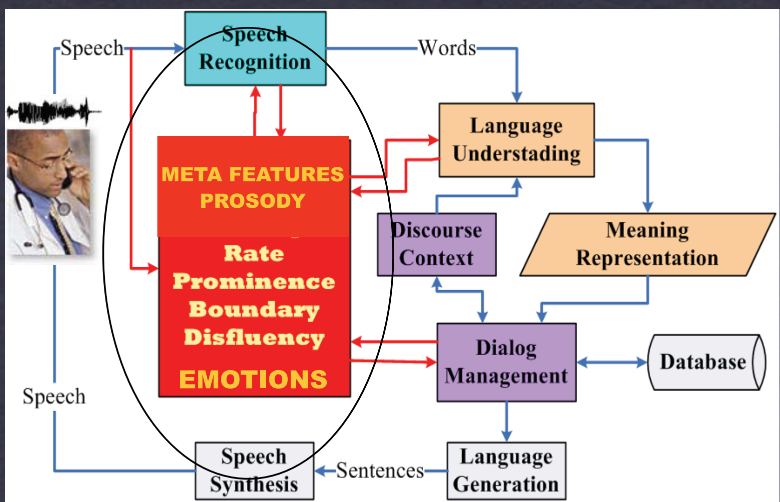
KNOWME NETWORKS: PERSONALIZED BEHAVIOR MEASUREMENTS IN REAL LIFE SETTINGS



PHYSICAL ACTIVITY, STRESS, COGNITIVE BEHAVIOR IN REALTIME

Enriched Speech & Language Processing: Is Multimodal!

mapping speech to words, and beyond



RECOGNIZE

- **WHAT**: SPOKEN LANGUAGE CONTENT
- **WHO**: SPEAKER IDENTITY,
- **HOW**: SPEAKING STYLE AND EMOTIONS

AUTOMATICALLY FROM SPEECH & SPOKEN LANGUAGE

Automatic Speech Recognition → So 'n' your chest pains have been going on just for two days is that right

What more can we infer beyond words?

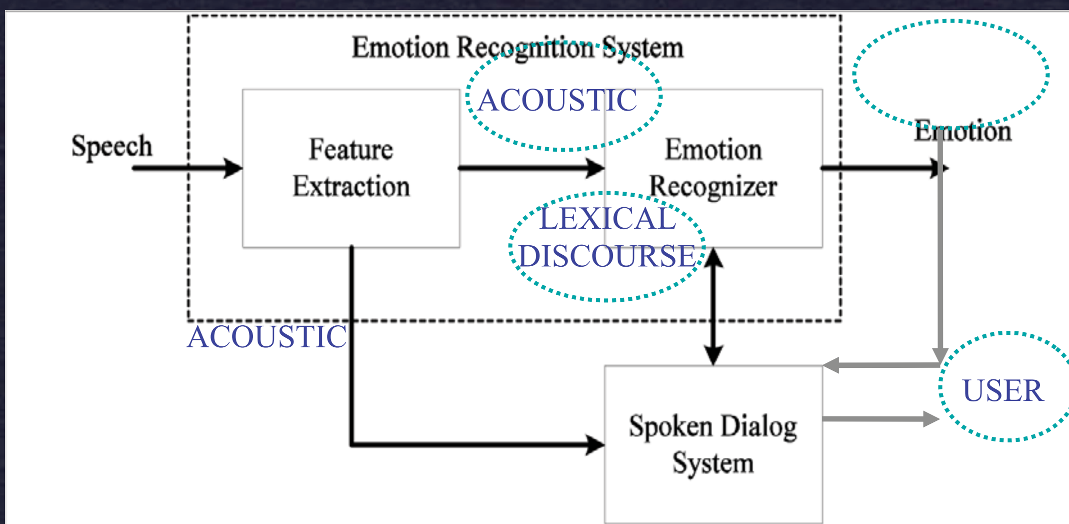
Words:	So 'n' your chest pains have been going on just for two days is that right
Speaker:	**spkr1 (Doctor)**
Prominent words:	So 'n' your **chest pains** have been going on just for **two days** is that **right**
Prosodic phrasing:	[So 'n' your chest] [pains have been going on just for two days] [is that right]
Speech act:	**Yes-No Question**
Affect:	**Neutral**
Attitude:	**Polite**

Rich information beyond words

23

Recognizing Emotions?

- **Expression versus experience**
- **Representations?**
  - Categorical (e.g., happy, sad..), Dimensional (arousal, valence)
  - Emotion Profiles to handle non protypical, blended emotions
  - Dynamic descriptions to capture changes in time



Lee & Narayanan, Toward recognizing emotions in spoken dialogs, IEEE Trans. Speech&Audio, 2005

24

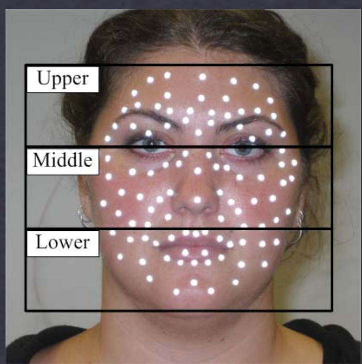
Enriching behavior descriptions further.... “Situated” Interactions & Conversational Computing

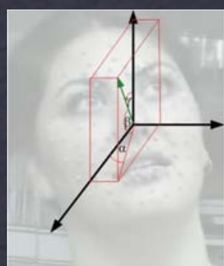
- Multimodality
- Interaction dynamics

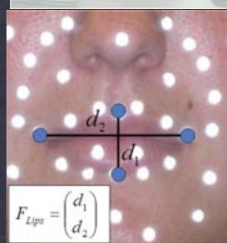
Modeling gestures/speech interrelation

VISUAL AND VOCAL FEATURES

- Speech
  - Prosodic features: Pitch, energy
  - MFCC coefficients (vocal tract)
- Visual features
  - Head motion
  - Eyebrow
  - Lips
  - Different face regions







C. Busso and S. Narayanan. Interrelation between Speech and Facial Gestures in Emotional Utterances: A single subject study. IEEE Transactions on Audio, Speech and Language Processing. 15(8): 2331 – 2347, November 2007.

Multimodal Emotion Recognition

- From speech
  - Average ~70%
  - Confusion sadness-neutral (□)
  - Confusion happiness-anger (□)
- From facial expression
  - Average ~85%
  - Confusion anger-sadness (□)
  - Confusion neutral-happiness (□)
  - Confusion sadness-neutral (□)
- Multimodal system (feature-level)
  - Average ~90%
  - Confusion neutral-sadness (□)
  - Other pairs are correctly separated

REDUNDANCY & COMPLEMENTARITY IN EMOTION ENCODING

USING SVM

	Anger	Sadness	Happiness	Neutral
Anger	0.68	0.05	0.21	0.05
Sadness	0.07	0.64	0.06	0.22
Happiness	0.19	0.04	0.70	0.08
Neutral	0.04	0.14	0.01	0.81

	Anger	Sadness	Happiness	Neutral
Anger	0.79	0.18	0.00	0.03
Sadness	0.06	0.81	0.00	0.13
Happiness	0.00	0.00	1.00	0.00
Neutral	0.00	0.04	0.15	0.81

	Anger	Sadness	Happiness	Neutral
Anger	0.95	0.00	0.03	0.03
Sadness	0.00	0.79	0.03	0.18
Happiness	0.02	0.00	0.91	0.08
Neutral	0.01	0.05	0.02	0.92

Busso et al, Analysis of emotion recognition using facial expressions, speech and multimodal information, ICMI, 2004

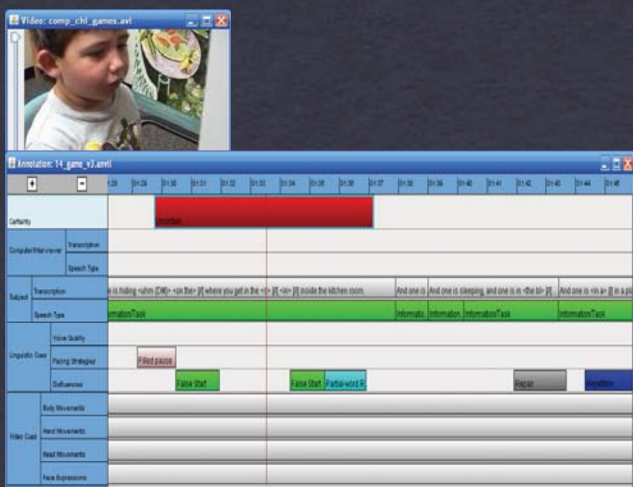
27

Is the child "certain"? (learning context)

TRAINED DECISION TREES (C4.5) USING "LEAVE-ONE-PERSON-OUT" TECHNIQUE

- USED DIFFERENT SUBSETS OF FEATURES: LEXICAL, ACOUSTIC, AND VISUAL
  - 6 LEXICAL/NONVERBAL VOCALIZATIONS: "I DON'T KNOW", FALSE STARTS, REPETITIONS, ETC.
  - 10 ACOUSTIC: PAUSES, ELONGATIONS, VOICE LOUDNESS, ETC.
  - 20 VISUAL: HEAD/FACE/HAND MOVEMENTS, FACIAL EXPRESSIONS

System	% Agreement *
Baseline (always guess certain)	72.32
Lexical	73.41
Visual	74.17
Acoustic	82.66
Lexical + Visual	74.17
Acoustic + Visual	81.94
Acoustic + Lexical	82.29
Acoustic + Lexical + Visual	82.66
Average Human Agreement	86.15



* ALL AGREEMENT STATISTICS ARE PAIRWISE AGREEMENT PERCENTAGES WITH THE GROUND TRUTH

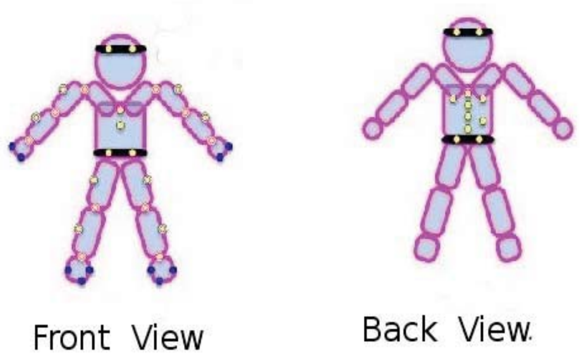
28

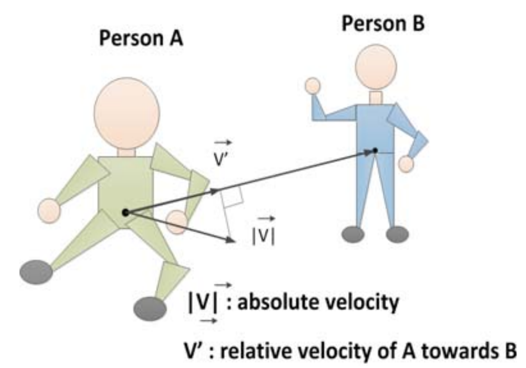
Tracking Emotion Trends using Body Language and Speech

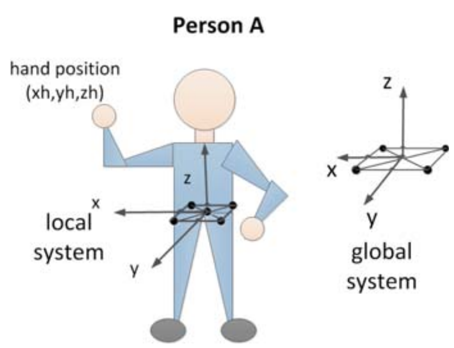
- Emotions continuously unfold with variable intensity and clarity
- Body language reflects the underlying emotion
- Goal
  - Track continuous emotional trends through time
    - Emotion tracking instead of emotion recognition
  - Understand how body language is modulated by emotion
- Approach/Contributions
  - Detailed body language descriptions
  - Joint modeling of audio-visual expressions and emotion
  - Tracking emotional trends at variable levels of detail
- Results/Findings
  - Promising performance for activation and dominance trend tracking
  - Importance of body language features
    - Looking at/away, leaning forward/away, walk vs run etc

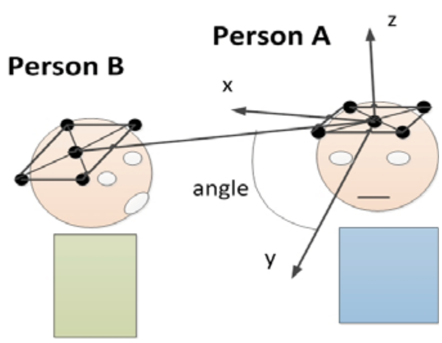
A. Metallinou, A. Katsamanis, Y. Wang and S. Narayanan, Tracking Changes in Continuous Emotion States using Body Language and Prosodic Cues, ICASSP 2011

Body Language Feature Extraction

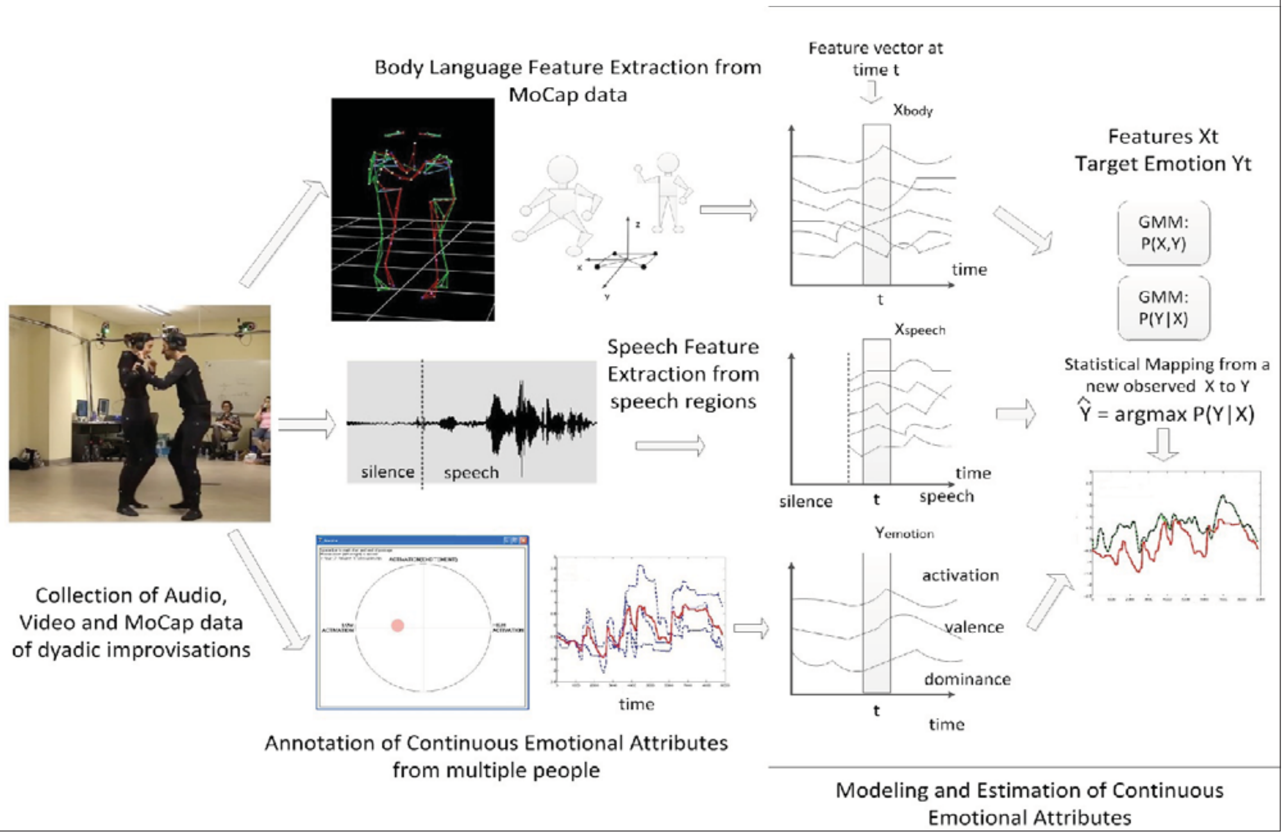








Framework Overview



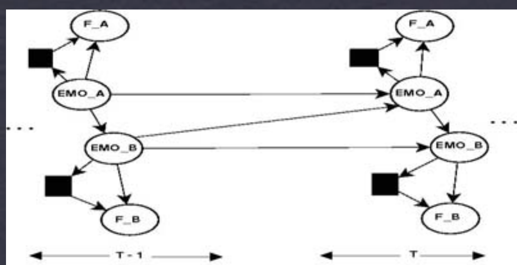
Emotion Tracking in Dyadic Spoken Interaction

- Problem
  - Emotion states tracking in dyadic spoken interaction
  - Incorporating "mutual influence" in the model
- Approach
  - Dynamic Bayesian Network: Joint modeling both speakers

•F0 Frequency
• Intensity/Energy
• Speech Rate

•Harmonic to Noise Ratio (HNR)
• 13 MFCC Coefficients
• 27 Mel Frequency Bank Filter Output

And functionals: Mean, Standard Deviation, Minimum, Maximum, 25% Quantile, 75% Quantile, Range, InterQuantile Range, Median, Kurtosis, Skewness



- Result
  - Emotion state tracking accuracy improves absolute 3.7%

Chi-Chun Lee, C. Busso, S. Lee and S. Narayanan, Modeling mutual influence of interlocutor emotion states in dyadic spoken interactions, in Proceedings of InterSpeech, 2009

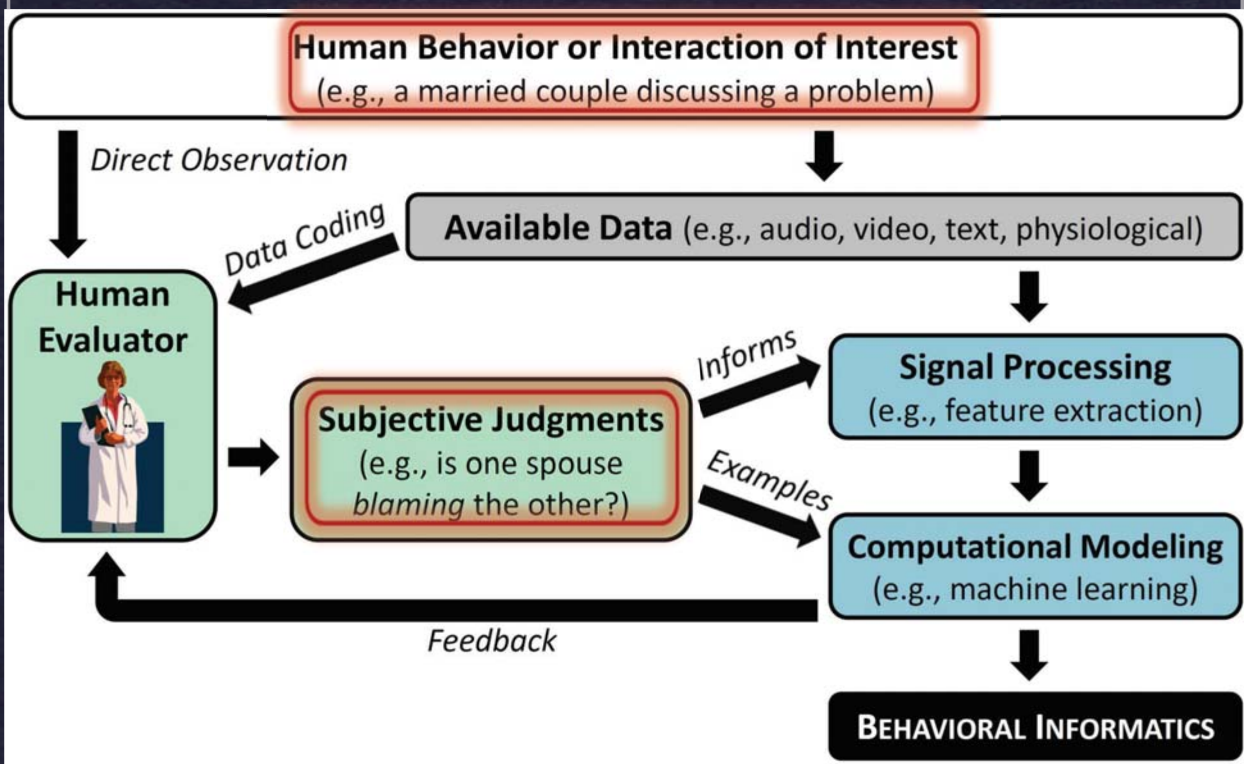
REST OF THIS TALK

Illustrative BSP applications

- *Couples Therapy: Behavior Coding*
- *Autism Spectrum Disorders*
- *Metabolic Health Monitoring*

33

BSP for Couples Therapy Research



34

Corpus

- **Real couples in 10-minute problem-solving interactions**
  - Longitudinal study at UCLA and U. of Washington [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
  - Video angle, microphone placement, and background noise varied

Human Behavior Coding

10-MINUTES LONG PROBLEM SOLVING INTERACTION

CODING IS PERFORMED AT THE SESSION-LEVEL

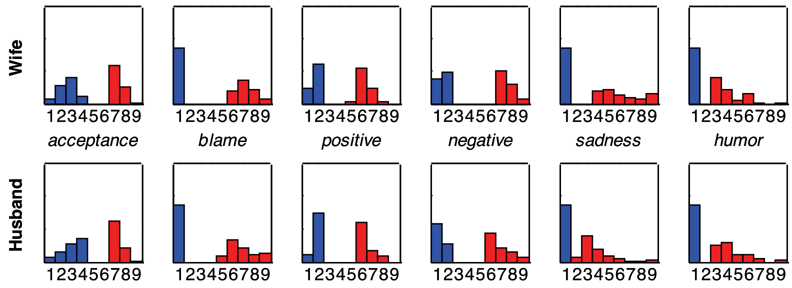
EXAMPLE CODING GOAL:
"IS THE HUSBAND SHOWING ACCEPTANCE?" (SCALE 1-9)

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

HUSBAND SPEAKING TURNS:

JONES AND CHRISTENSEN, SOCIAL SUPPORT INTERACTION RATING SYSTEM, UCLA, 1998

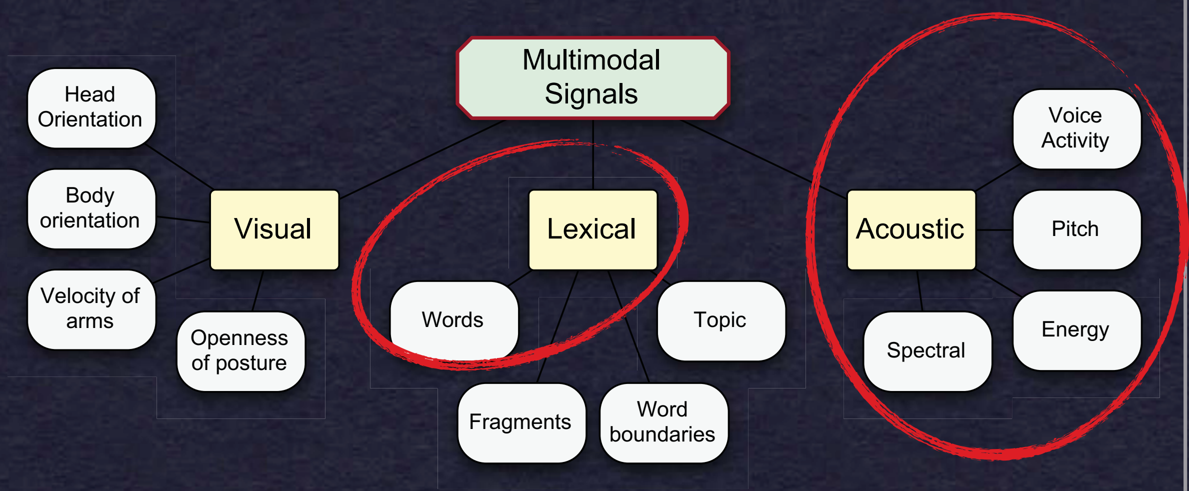
Focus on extreme cases of session-level judgments



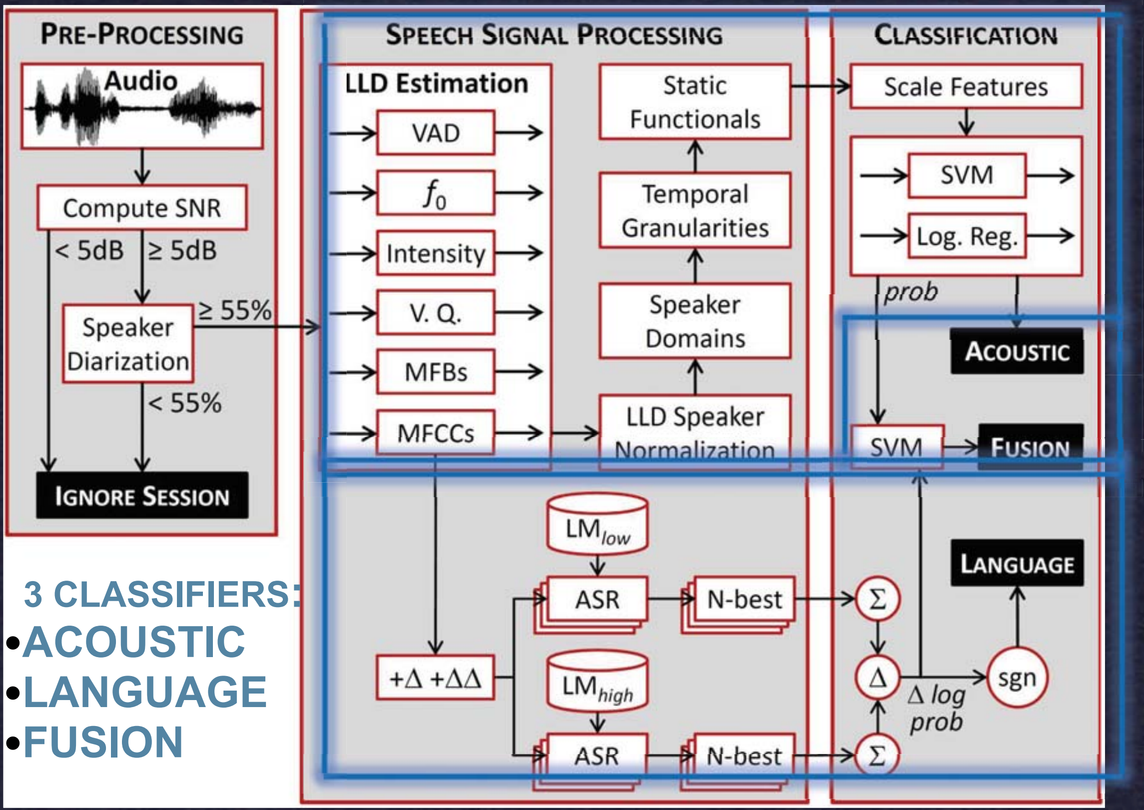
- **Initial work focused on 6 codes:**
  - acceptance, blame, positive affect, negative affect, sadness, humor

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. 2011

Estimate behavioral codes from audio information?



Methodology Overview

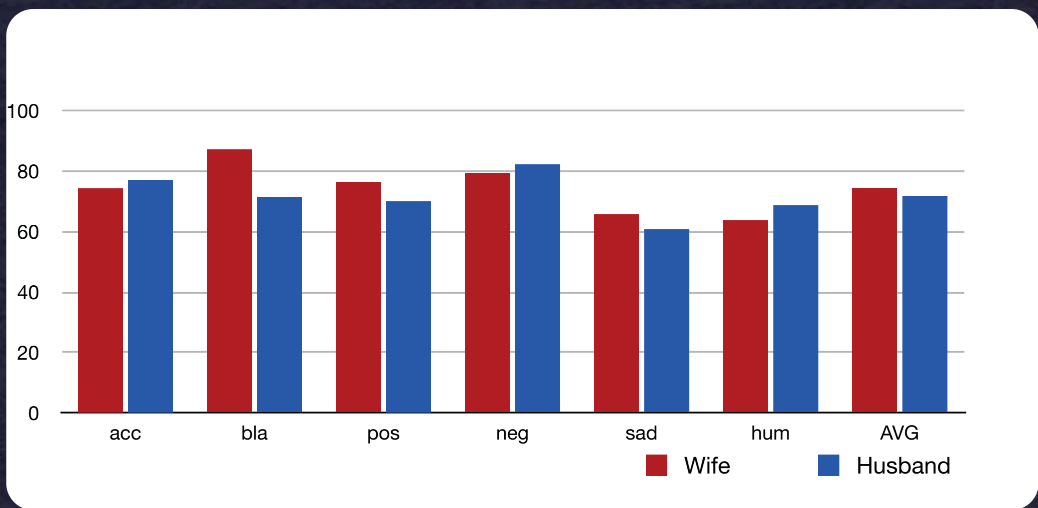


3 CLASSIFIERS:
- ACOUSTIC
- LANGUAGE
- FUSION

39

(Very) Simple Acoustic-feature based Behavior Estimation

- Use of acoustic low-level descriptors (LLDs)
  - Binary classification task
  - Linear-SVM (best so far)
  - Global speaker-dependent cues capture evaluators' codes well



M. BLACK, ET AL "AUTOMATIC CLASSIFICATION OF MARRIED COUPLES' BEHAVIOR USING AUDIO FEATURES" INTERSPEECH 2010

Lexical-information based Behavior Estimation

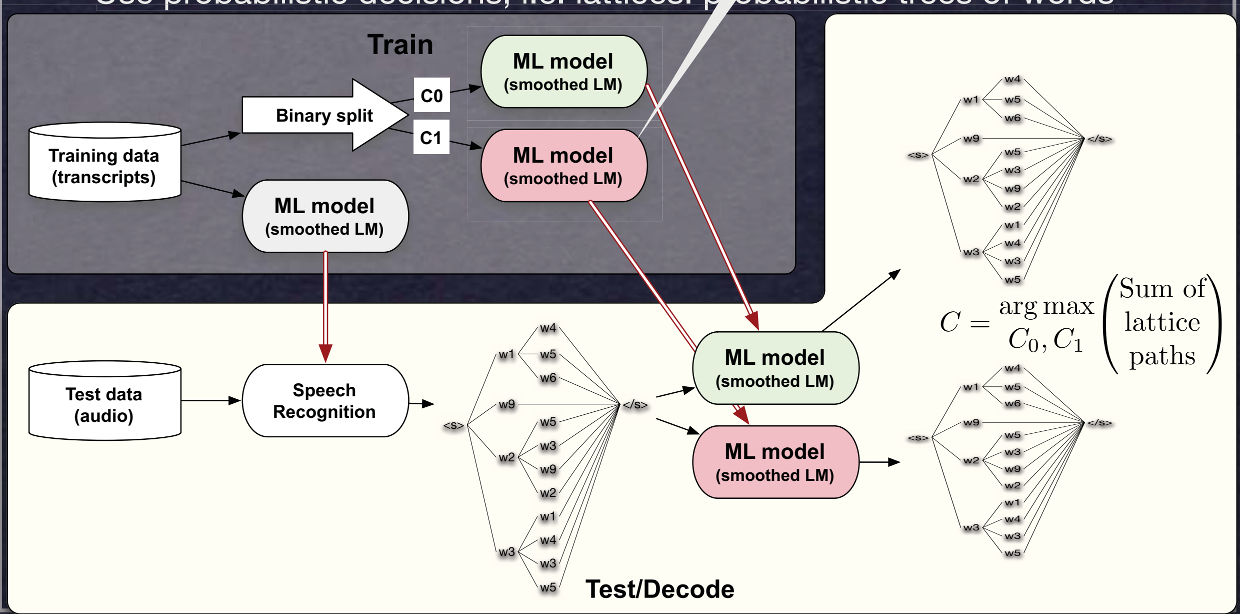
Partner	Transcript
H	WHAT DID I TELL YOU YOU CAN DO THAT AH AND EVERYTHING
W	BUT WHY DID YOU ASK THEN WHY DID TO ASK
H	AND DO IT MORE AND GET US INTO TROUBLE
W	YEAH WHY DID YOU ASK SEE MY QUESTION IS
H	MM HMMM
W	IF IF YOU TOLD ME THIS AND I AGREE I WOULD KEEP TRACK OF IT AND EVERYTHING
H	THAT’S THAT’S
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

Lexical Behavior Estimation

- Logical way is to transcribe
- Speech recognition is not perfect
- Use probabilistic decisions, i.e. lattices: probabilistic trees of words

Context: Data mining can enrich these



Informing experts

- Through the lexical analysis we can inform experts
  - Example: Top and bottom words that contributed to (correct) classification of a partner as “blaming”

High Blame		Low Blame	
word	Δ log	word	Δ log
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

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

43

Some challenges (and preliminary) 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 dependency between interlocutors
  - ➡ **Models of entrainment**
- Not all human observers/evaluators are equally reliable, and reliability is data dependent
  - ➡ **Realistic models of human observers/evaluators**

44

Fusion Results: Estimating "Blame"

– Exploit complementary information from language and speech
– Score-level fusion of classifiers using confidence scores

Classifier Type	Accuracy
Baseline Chance	50.0%
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

Behavior Collection Space: Multichannel Multimodal

Audio:

- 3 4-mic T-arrays
- 2 lapel mics
- 1 shotgun mic

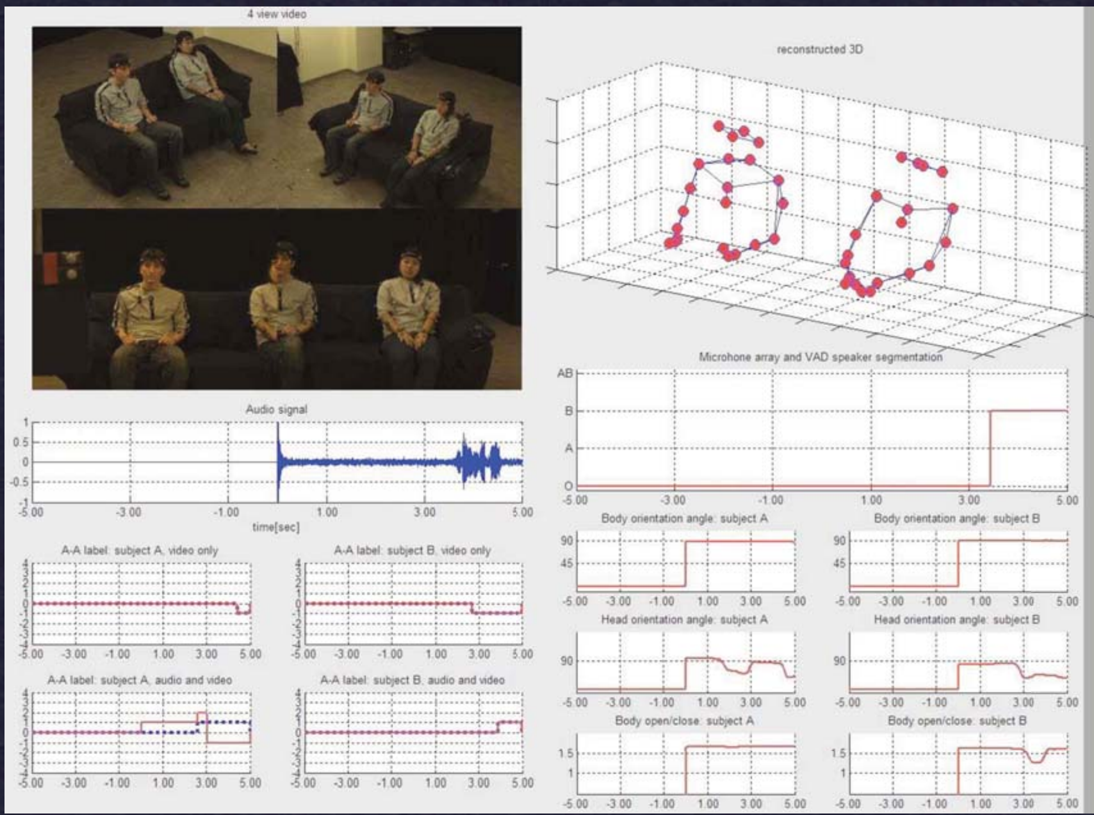
Video:

- 10 HD cameras (PointGrey Flea 2)
- Motion capture: 12 ViconQ Sensors

Accurate synchronization

V. ROZGIĆ, B. XIAO, A. KATSAMANIS, B. BAUCOM, P. G. GEORGIOU, AND S. NARAYANAN, "A NEW MULTICHANNEL MULTIMODAL DYADIC INTERACTION DATABASE" INTERSPEECH 2010

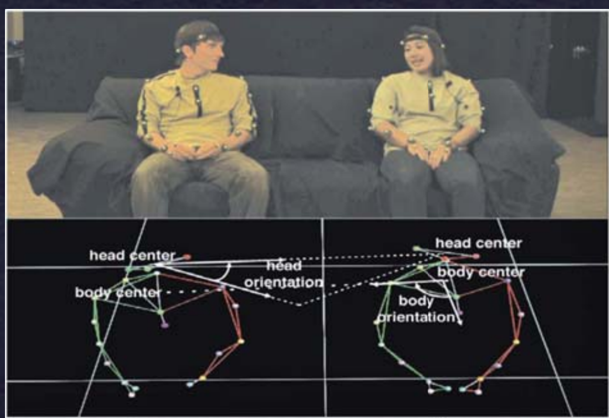
Example Multimodal Data



47

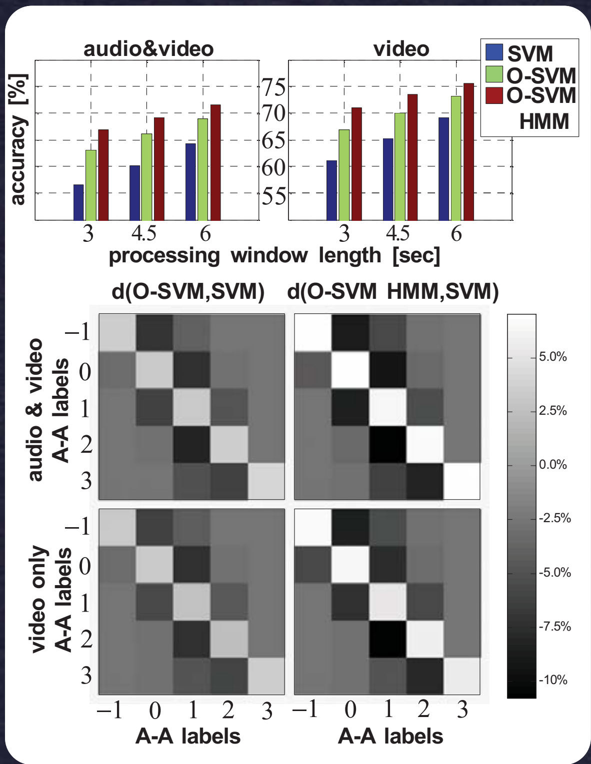
Low Level Descriptor Features

- Audio features:
  - pitch
  - energy
  - MFCCs
  - speaker segmentation using VAD/Mic array
- Motion capture features:
  - Head/body orientation relative to the other subject
  - Arm velocity maximized over left and right hand
  - Body open/close in terms of average distance of left and right forearms to chest
- Functionals: mean, min, max, std of features on 3sec intervals
  - (3s windows with 1s overlap are motivated by the **3s coding rule**)



48

APPROACH-AVOIDANCE (A-A) CODE ESTIMATION RESULTS



- Reference codes from experts:
  - USING AUDIO & VIDEO, OR ONLY VIDEO
- Three methods are compared:
  - PLAIN MULTI-CLASS SVM
  - ORDINAL SVM
  - ORDINAL SVM WITH HMM SMOOTHING

ON THE ESTIMATED LABEL SEQUENCE

- Ordinal SVM takes inherent order information of A-A codes into account
- Difference of confusion matrices show the advantage of ordinal regression method

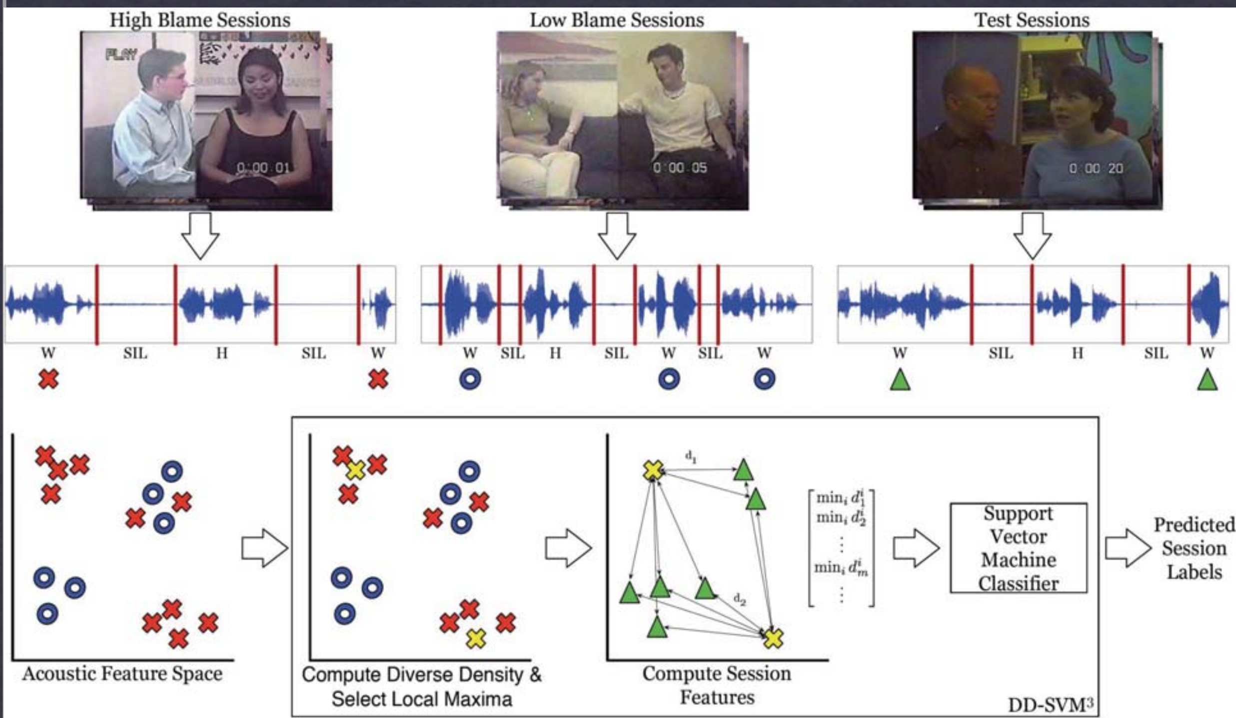
V. ROZGIĆ, B. XIAO, A. KATSAMANIS, B. BAUCOM, P. G. GEORGIOU, AND S. NARAYANAN. ESTIMATION OF ORDINAL APPROACH-AVOIDANCE LABELS IN DYADIC INTERACTIONS: ORDINAL LOGISTIC REGRESSION APPROACH. IN PROCEEDINGS OF ICASSP 2011

Some challenges (and preliminary) approaches

- Any single feature stream offers partial, noisy code information
  - ➡ **Multimodal approach**
- Behavior ratings are relative, often on an ordered scale
  - ➡ **Ordinal regression**
- ✓ Not all portions of the feature stream are equally relevant in explaining an overall behavior description
  - ➡ **Salient instances: Multiple instance learning**
- Not all human observers/evaluators are equally reliable, and reliability is data dependent
  - ➡ **Realistic models of human observers/evaluators**
- Behavior is a part of an interaction: mutual dependency between interlocutors
  - ➡ **Models of entrainment**

Saliency Detection with Multiple Instance Learning and Diverse Density SVM

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., "AUTOMATIC IDENTIFICATION OF SALIENT ACOUSTIC INSTANCES IN COUPLES' BEHAVIORAL INTERACTIONS USING DIVERSE DENSITY SVM," IN INTERSPEECH 2011.

Summary: Some challenges (and preliminary) approaches

- ✓ Any single feature stream offers partial, noisy code information
  - ➡ **Multimodal approach**
- ✓ Behavior ratings are relative, often on an ordered scale
  - ➡ **Ordinal regression**
- ✓ Not all portions of the feature stream are equally relevant in explaining an overall behavior description
  - ➡ **Salient instances: Multiple instance learning**
- ✓ Behavior is a part of an interaction: mutual dependency between interlocutors
  - ➡ **Models of entrainment**
- ✓ Not all human observers/evaluators are equally reliable, and reliability is data dependent
  - ➡ **Realistic models of human observers/evaluators**

REST OF THIS TALK

- Some BSP building blocks

Example Behavioral Analysis Studies

- Family Studies: Marital couples
  - Blame patterns; positiveness/negativeness; humor/sarcasm
- Autism Spectrum Disorders
  - Characterizing and quantifying socio-emotional discourse
  - Technology interfaces for elicitation & personalized interventions
- Metabolic Health Monitoring
  - Technologies for ecologically valid measurements
  - Novel just in time analytics and intervention technology support

53

Autism Spectrum Disorders (ASD)

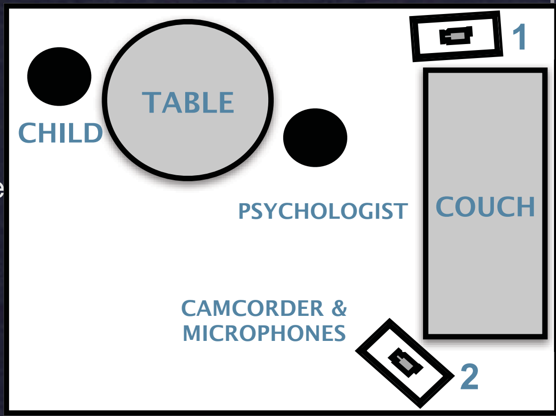
- **1 in 110 children diagnosed with ASD (CDC, 2010)**
- **ASD characterized by**
  - Difficulties in social communication, reciprocity
  - Repetitive or stereotyped behaviors and interests

Technology possibilities in ASD include

- Computational techniques to
  - Better understand communication and social patterns of children
  - Help stratify phenotyping with quantifiable and adaptable metrics
  - Automatically track children's progress during interventions
- Interfaces/systems to elicit, encourage, analyze behavior:
  - Complex, but phased; Structured; Naturalistic

USC CARE Corpus

- **Child-psychologist interactions of children with autism in the context of Autism Diagnostic Observation Schedule (ADOS)**
  - ADOS is a widely used clinical research instrument, comprised of semi-structured social activities to provide psychologist with sample of behavior (30-60 minutes)
  - Psychologist then rates child on autism-relevant symptoms and provides final classification of autism/ASD
- **USC CARE Corpus**
  - *CARE*: Center for Autism Research in Engr.
  - *Collaboration*: USC Keck School of Medicine
  - Recorded ADOS of 75+ children (goal = 100)
  - Controlled, clinical environment
  - 2 HD camcorders + 2 far-field microphones
  - Access to psychologists' notes and codes

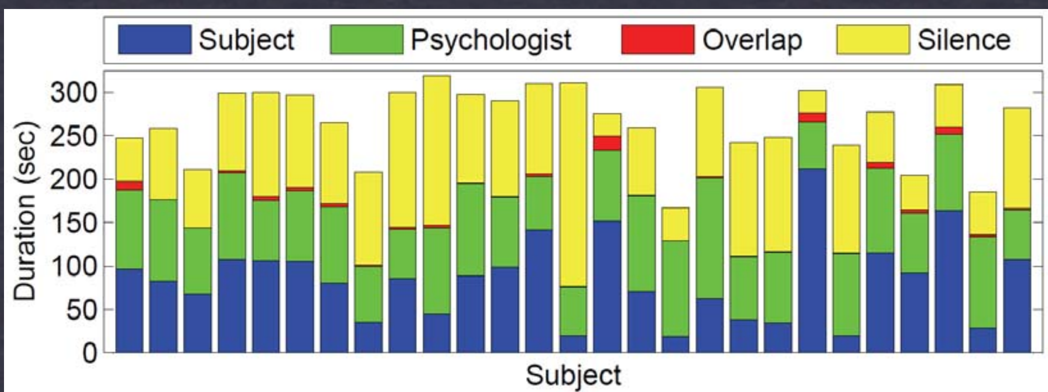


M. Black, D. Bone, M. Williams, P. Gorrindo, P. Levitt, and S. Narayanan. The USC CARE Corpus: Child-Psychologist Interactions of Children with Autism Spectrum Disorders. In Proceedings of InterSpeech, 2011.

55

Analysis: Speech Activity

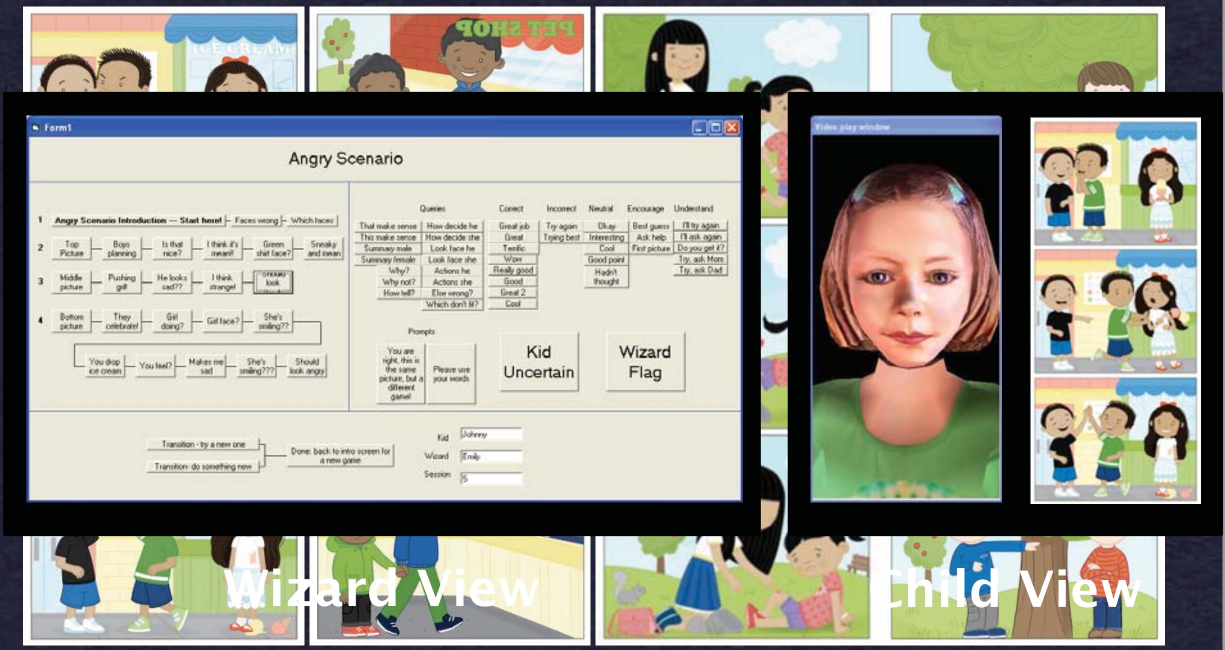
- COMPUTED SPEECH/NON-SPEECH STATISTICS



- **Duration of subject's speech correlated with 14/33 ADOS codes & totals**
  - *P* < 0.001: OFFERS INFORMATION, REPORTING OF EVENTS, INSIGHT
  - *P* < 0.01: EMPATHY/COMMENTS ON OTHER'S EMOTIONS, COMMUNICATION TOTAL, COMMUNICATION AND SOCIAL INTERACTION TOTAL
  - *P* < 0.05: IMMEDIATE ECHOLIA, CONVERSATION, SHARED ENJOYMENT, AMOUNT OF RECIPROCAL SOCIAL COMMUNICATION, IMAGINATION/CREATIVITY, SOCIAL INTERACTION TOTAL, IMAGINATION/CREATIVITY TOTAL, ADOS CLASSIFICATION

56

Interactions with computer characters: Emotional Reasoning Games:

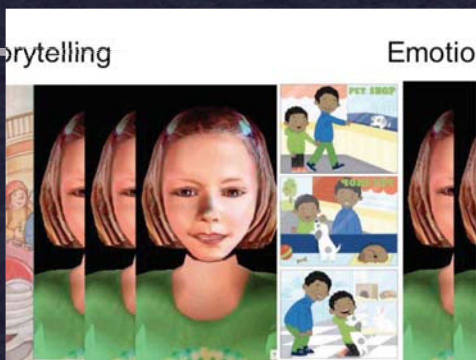


IN-ORDER | OUT-OF-ORDER | MISSING FACES | MISMATCHED FACES

CSLU TOOLKIT

Data

- **Automatic data logging**
  - ECA behavior
  - Wizard flag
- **Recorded data: Clinic**
  - Three Sony Handycams
  - Two shotgun microphones
- **Extracted data**
  - Manual transcription – Word usage statistics
  - Audio feature extraction



EMILY MOWER, MATTHEW BLACK, ELISA FLORES, MARIAN WILLIAMS AND SHRIKANTH NARAYANAN. DESIGN OF AN EMOTIONALLY TARGETED INTERACTIVE AGENT FOR CHILDREN WITH AUTISM. IN PROCEEDINGS OF ICME 2011

Initial Findings: Remarks

- **Given the experimental setup, it is practical to extract:**
  - Prosodic information (pitch, energy of speech)
  - Spectral information (measure of frequency content of speech)
  - Word Usage statistics (e.g., number of backchannels)
- **The results demonstrate that in general:**
  - The parent's audio data provides discriminatory information regarding the ECA presence or absence
  - The child's audio data does not

Suggests: similarity in the child's communicative behavior across the two interaction conditions, with and without the ECA present

Emily Mower, Chi-Chun Lee, James Gibson, Theodora Chaspari, Marian Williams, Shrikanth Narayanan. Analyzing the Nature of ECA Interactions in Children with Autism.. In Proceedings of InterSpeech, 2011.

THIS TALK

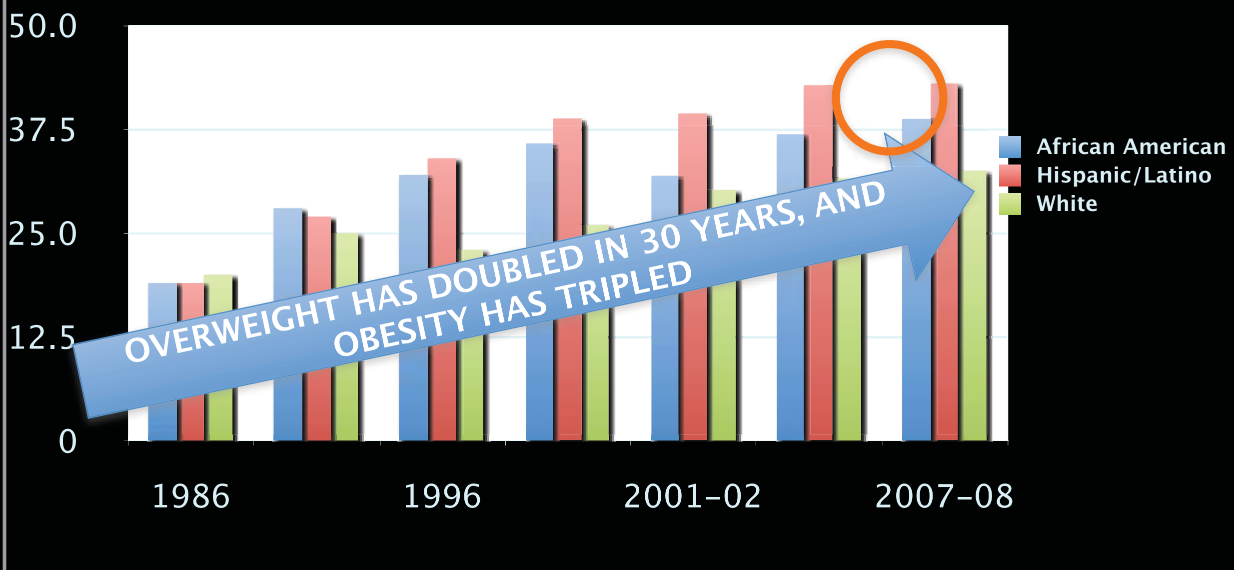
- Some BSP building blocks

Example Behavioral Analysis Studies

- Family Studies: Marital couples
  - Blame patterns; positiveness/negativeness; humor/sarcasm
- Autism Spectrum Disorders
  - Characterizing joint attention; quantifying socio-emotional discourse
  - Technology interfaces for elicitation and personalized interventions
- Metabolic Health Monitoring
  - Characterizing physical behavior in context

60

BMI ≥ 85th Percentile (age 4-19)



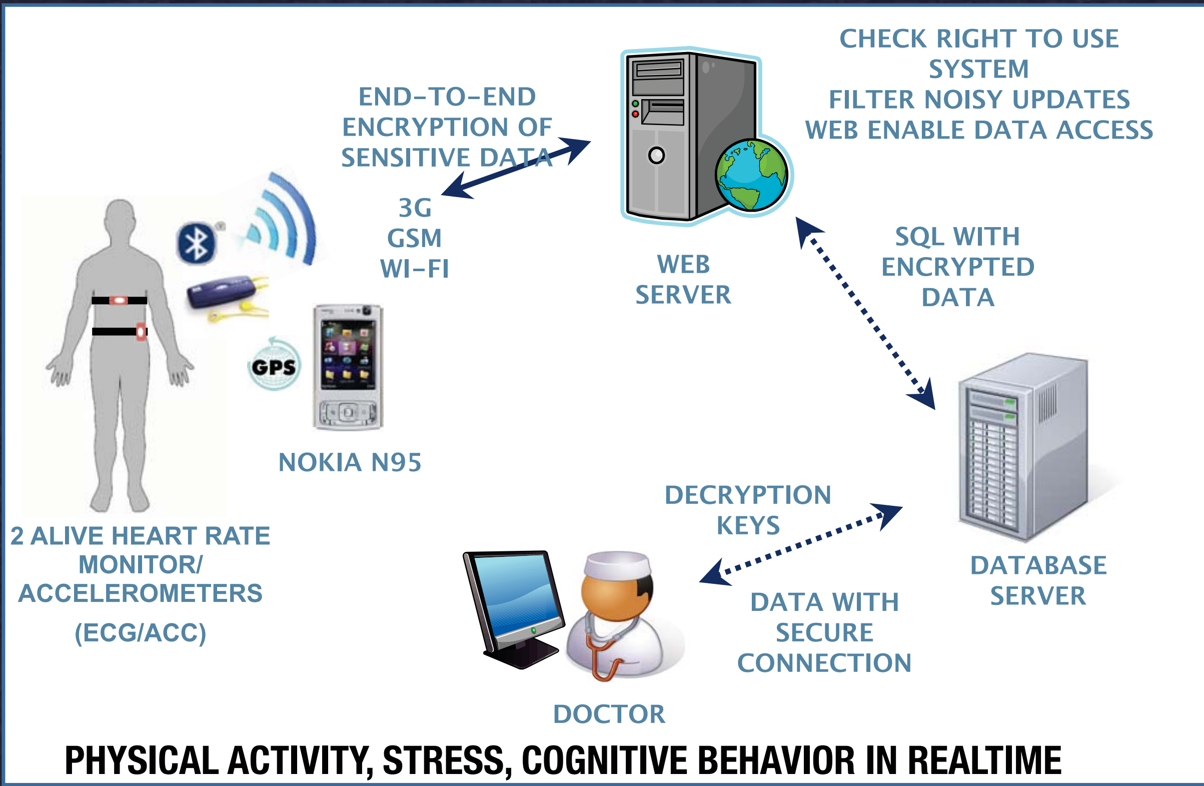
➢ ***PERSISTS INTO ADULTHOOD (WHITAKER ET AL. NEJM: 1997;337:869-873)***

DATA FROM THE NATIONAL LONGITUDINAL SURVEY OF YOUTH 1986-1998, NHANES 1998-2008

KNOWME Networks

- A suite of mobile, Bluetooth-enabled, wireless, wearable sensors
- That interface with a mobile phone and secure server
- To process data in real time,
- Designed specifically for use in overweight minority youth

KNOWME Networks: Current System



KNOWME NETWORKS

IN-LAB & FIELD DEVELOPMENT: BEHAVIORAL SIGNAL PROCESSING

ADAR EMKEN, MING LI, GAUTAM THATTE, SANGWON LEE, MURALI ANNAVARAM, URBASHI MITRA, SHRIKANTH NARAYANAN, AND DONNA SPRUIJT-METZ. RECOGNITION OF PHYSICAL ACTIVITIES IN OVERWEIGHT HISPANIC YOUTH USING KNOWME NETWORKS. JOURNAL OF PHYSICAL ACTIVITY & HEALTH. 2011.

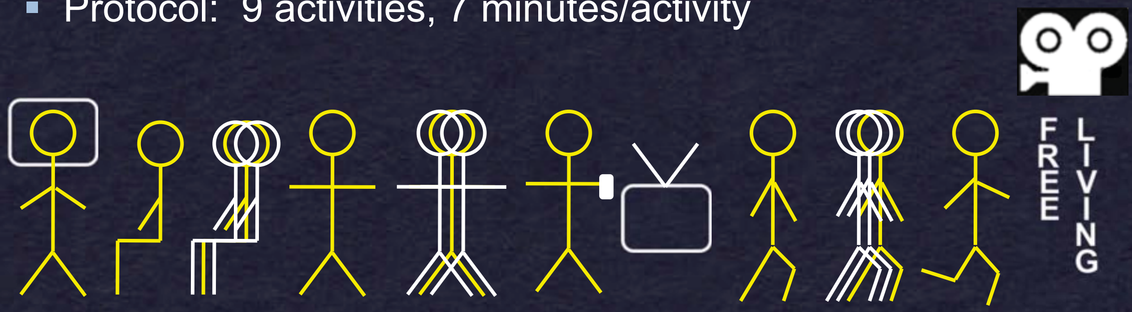
GAUTAM THATTE, MING LI, SANGWON LEE, ADAR EMKEN, MURALI ANNAVARAM, SHRIKANTH NARAYANAN, DONNA SPRUIJT-METZ, AND URBASHI MITRA. OPTIMAL TIME-RESOURCE ALLOCATION FOR ENERGY-EFFICIENT PHYSICAL ACTIVITY DETECTION. IEEE TRANSACTIONS ON SIGNAL PROCESSING. 2011.

GAUTAM THATTE, MING LI, SANGWON LEE, ADAR EMKEN, SHRIKANTH NARAYANAN, URBASHI MITRA, DONNA SPRUIJT-METZ AND MURALI ANNAVARAM. KNOWME: AN ENERGY-EFFICIENT, MULTIMODAL BODY AREA NETWORK FOR PHYSICAL ACTIVITY MONITORING. ACM TRANSACTIONS ON EMBEDDED COMPUTING SYSTEMS. 2011.

MING LI, VIKTOR ROZGIC, GAUTAM THATTE, SANGWON LEE, ADAR EMKEN, MURALI ANNAVARAM, URBASHI MITRA, DONNA SPRUIJT-METZ AND SHRIKANTH NARAYANAN. MULTIMODAL PHYSICAL ACTIVITY RECOGNITION BY FUSING TEMPORAL AND CEPSTRAL INFORMATION. IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING. 18(4): 369-380, 2010.

In-Lab Physical Activity Detection

- Developed and tested algorithms to classify PA types in 20 overweight Hispanic youth
  - 10 F/10M; 14.6 ± 1.8 years old; BMI %tile 96 ± 4
- Protocol: 9 activities, 7 minutes/activity



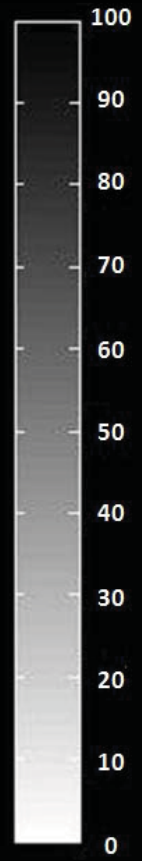
- 3 sessions develop algorithm/1 session test algorithm

PREDICTED BY THE MODEL (84-94% ACCURACY)

% NORMALIZED ACROSS EACH ROW

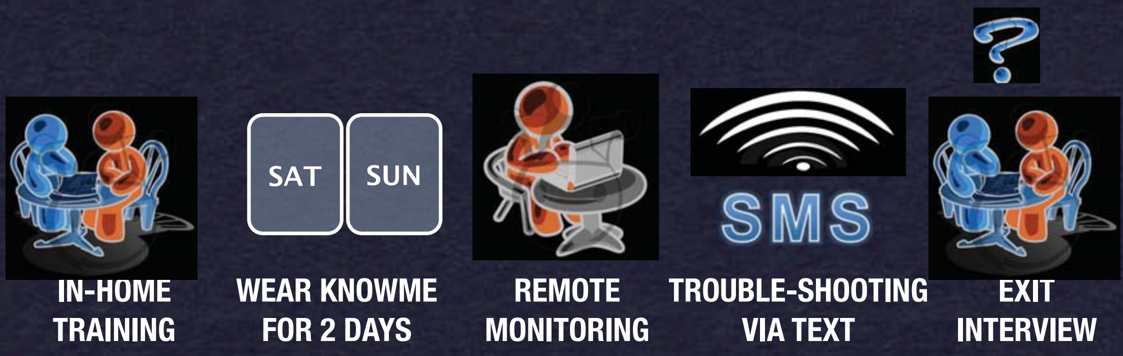
Actual

1270	50	24	24	0	6	2	7	18
128	788	372	77	4	13	2	0	1
46	395	781	89	72	17	5	2	0
44	33	30	1020	245	22	0	0	0
3	0	23	162	1012	139	19	0	0
0	0	0	36	52	1325	13	0	0
0	0	0	0	30	7	1335	58	0
0	0	0	0	2	2	104	1297	25
0	0	0	0	0	0	0	38	1102

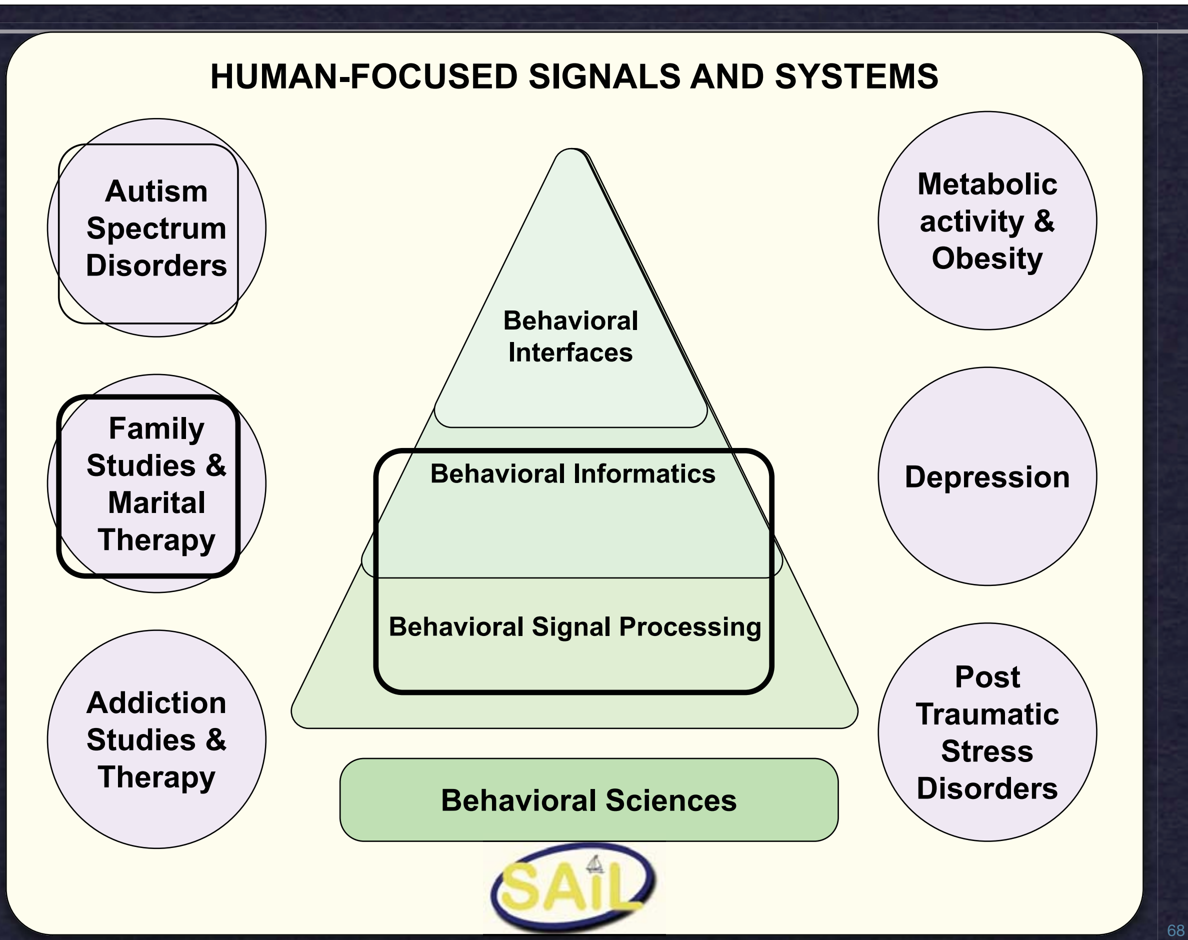


Will/Can youth use this system in the real?

- **Subjects: 12 overweight Hispanic youth**
  - 5F/7M; 14.8 ± 1.9 years old; BMI %tile 97 ± 3
- **Protocol:**



- **Results/day**
  - Wore KNOWME for 11.4 ± 2.0 hours
  - Phone battery life 9.2 ± 2.6 hours
  - 8 SMS sent to us / 9 SMS received from us



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 scientific and societal possibilities**
  - Opportunities for interdisciplinary and collaborative scholarship

BEHAVIORAL SIGNAL PROCESSING:

✓HELP DO THINGS WE KNOW TO DO WELL MORE EFFICIENTLY, CONSISTENTLY
✓HELP HANDLE NEW DATA, CREATE NEW MODELS TO OFFER UNIMAGINED INSIGHTS

SAiL

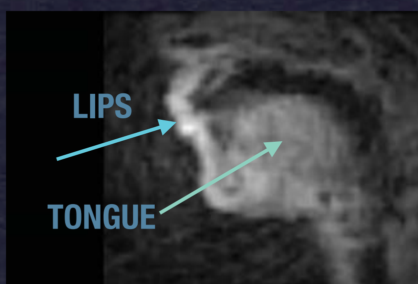
USC Viterbi School of Engineering

http://sail.usc.edu/

WORK REPORTED REPRESENTS COLLABORATIVE EFFORTS WITH NUMEROUS COLLEAGUES AND COLLABORATORS

SUPPORTED BY: ONR, ARMY, DARPA, NSF AND NIH

REALTIME MRI OF VOCAL PRODUCTION



REALTIME VOCAL TRACT MRI OF BEATBOXING