USC University of Southern California

Behavioral Signal Processing:
Enabling human-centered behavioral informatics

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

University of Southern California

MIT -CSAIL
February 2015

USC University of Southern California

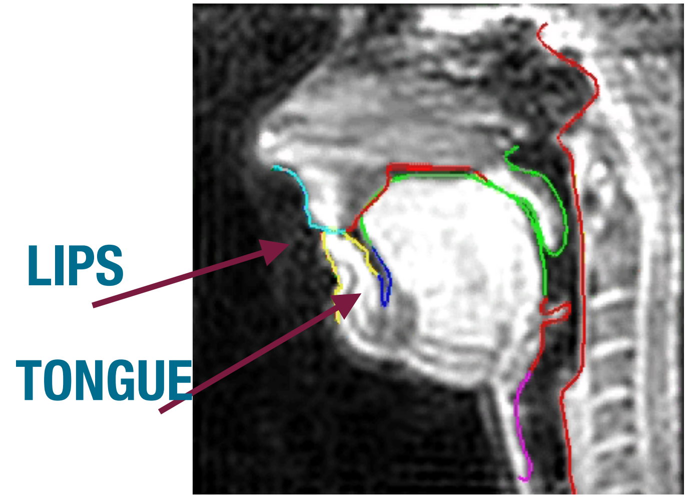
SAIL

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

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

Prof. Ken Stevens, 1924-2013

To whom we owe a lot...



23 frames/sec

REALTIME MRI OF VOCAL PRODUCTION

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

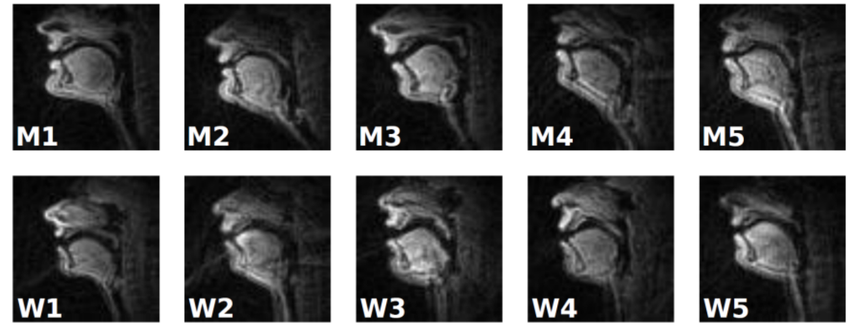
Current protocol: 96 frames/second

Sajan Lingala, Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan, Krishna Nayak. A fast and flexible MRI system for the study of dynamic vocal tract shaping. Magnetic Resonance in Medicine. 2016

42 frames/sec: 2 planes simultaneously

USC-TIMIT: A MULTIMODAL ARTICULATORY DATA CORPUS FOR SPEECH RESEARCH

- 10 American English talkers (5M, 5F).
- Real time MRI (5 speakers also with EMA) and synchronized audio.
- 460 sentences each (>20 minutes)
- Freely available for speech research.



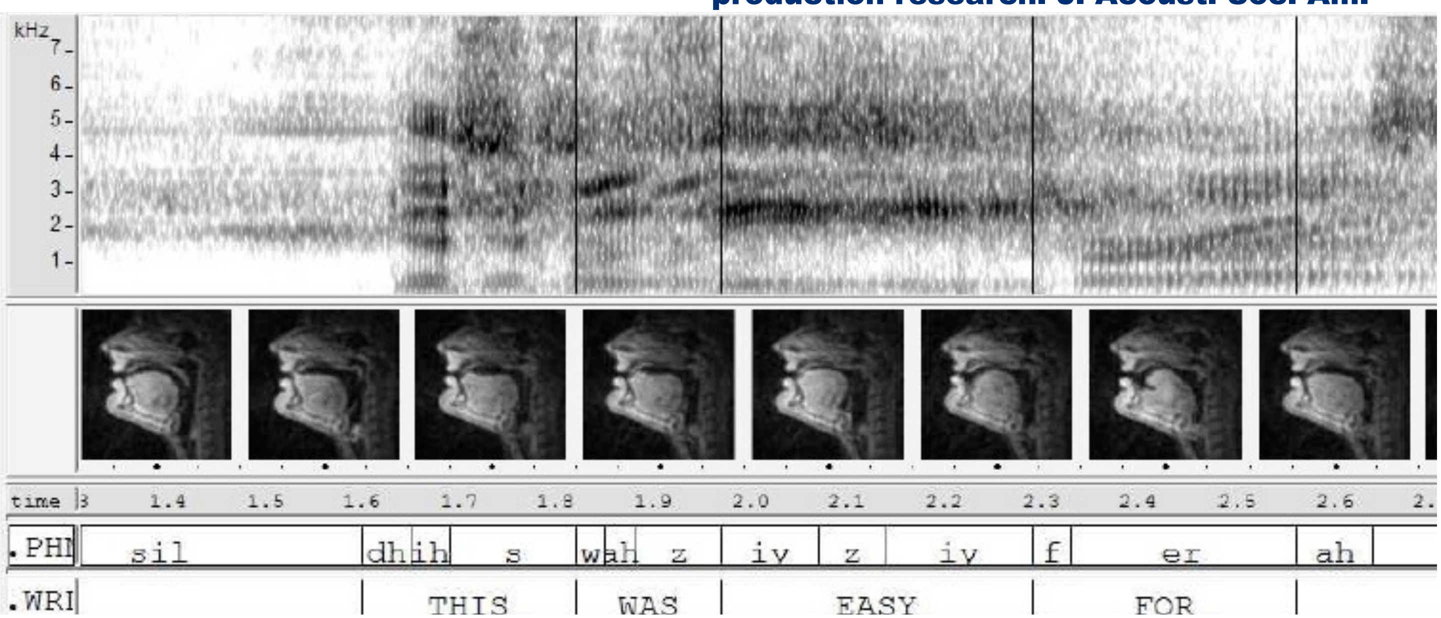
WEB-LINK (with download info):

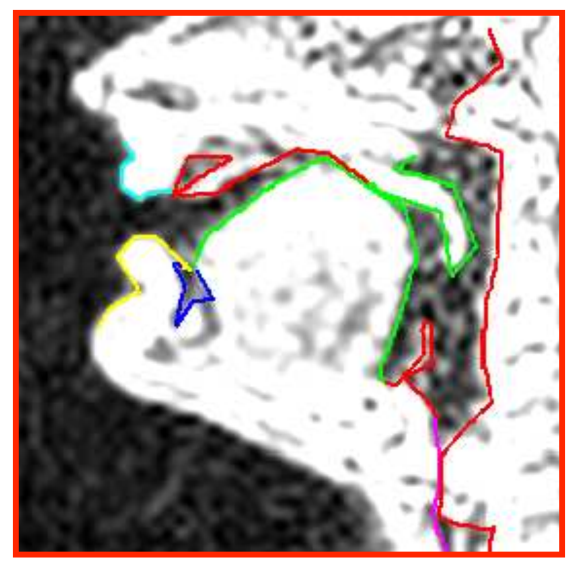
http://sail.usc.edu/span/usc-timit/

SAIL homepage: http://sail.usc.edu

Narayanan et al. (2011). A Multimodal Real-Time MRI Articulatory Corpus for Speech Research. InterSpeech.

Narayanan et al. (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research. J. Acoust. Soc. Am.





Seeking a window into the human mental state

through engineering approaches

BEHAVIORAL SIGNAL PROCESSING:

COMPUTING BEHAVIORAL TRAITS & STATES FOR DECISION MAKING AND ACTION

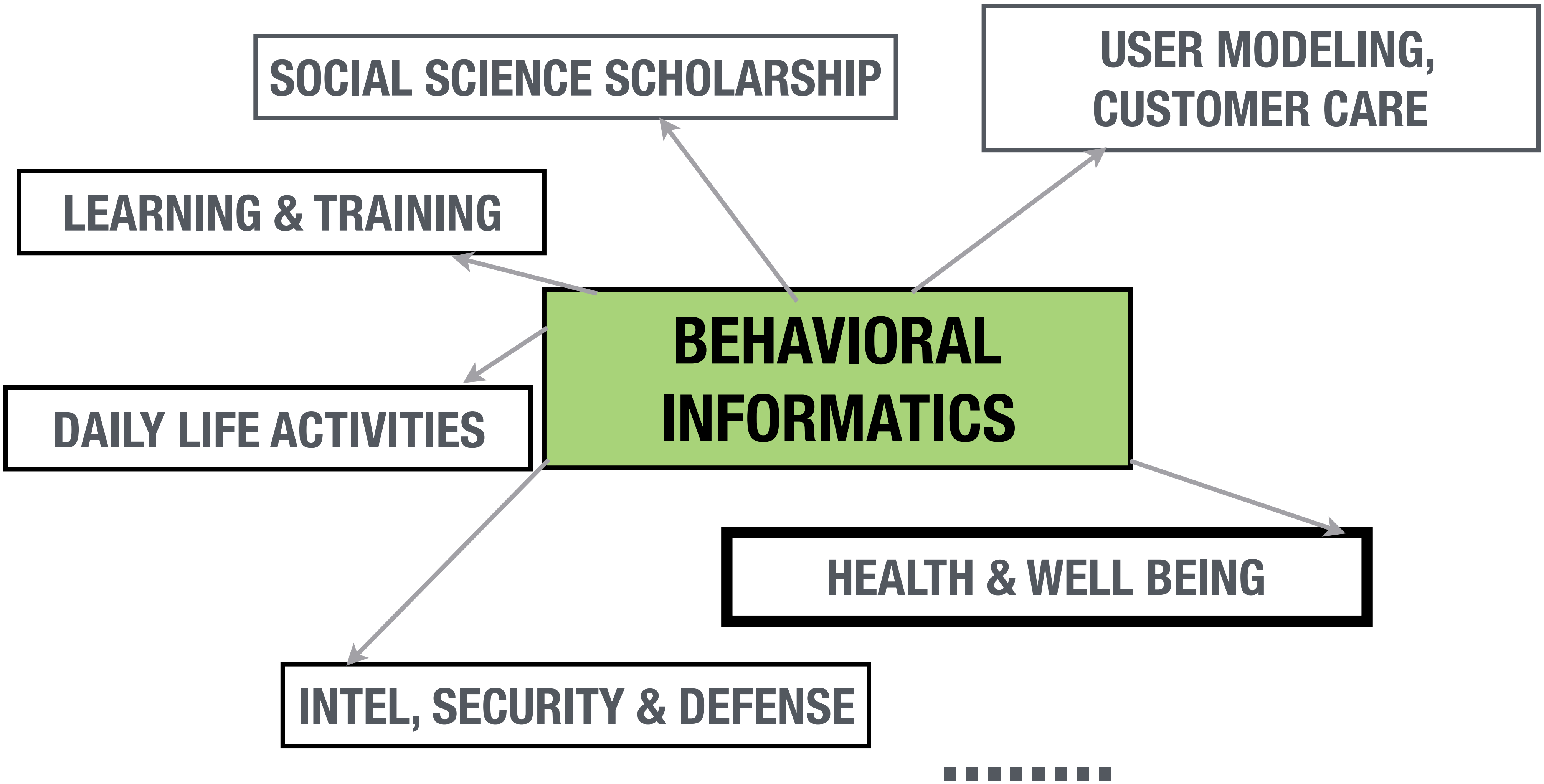
✓ HELP DO THINGS WE KNOW TO DO WELL MORE EFFICIENTLY, CONSISTENTLY

✓ HELP HANDLE NEW DATA, CREATE NEW MODELS TO OFFER UNIMAGINED INSIGHTS: CREATE TOOLS FOR DISCOVERY

- FOCUS OF THE TALK ON SPEECH AND SPOKEN LANGUAGE CUES
- HEALTH & WELL BEING APPLICATIONS

BEHAVIOR ANALYSIS CENTRAL TO MANY ENDEAVORS

..BOTH IN BASIC RESEARCH AND ACROSS APPLICATION DOMAINS



ROLE OF ENGINEERING?

Many facets & perspectives: “informatics”

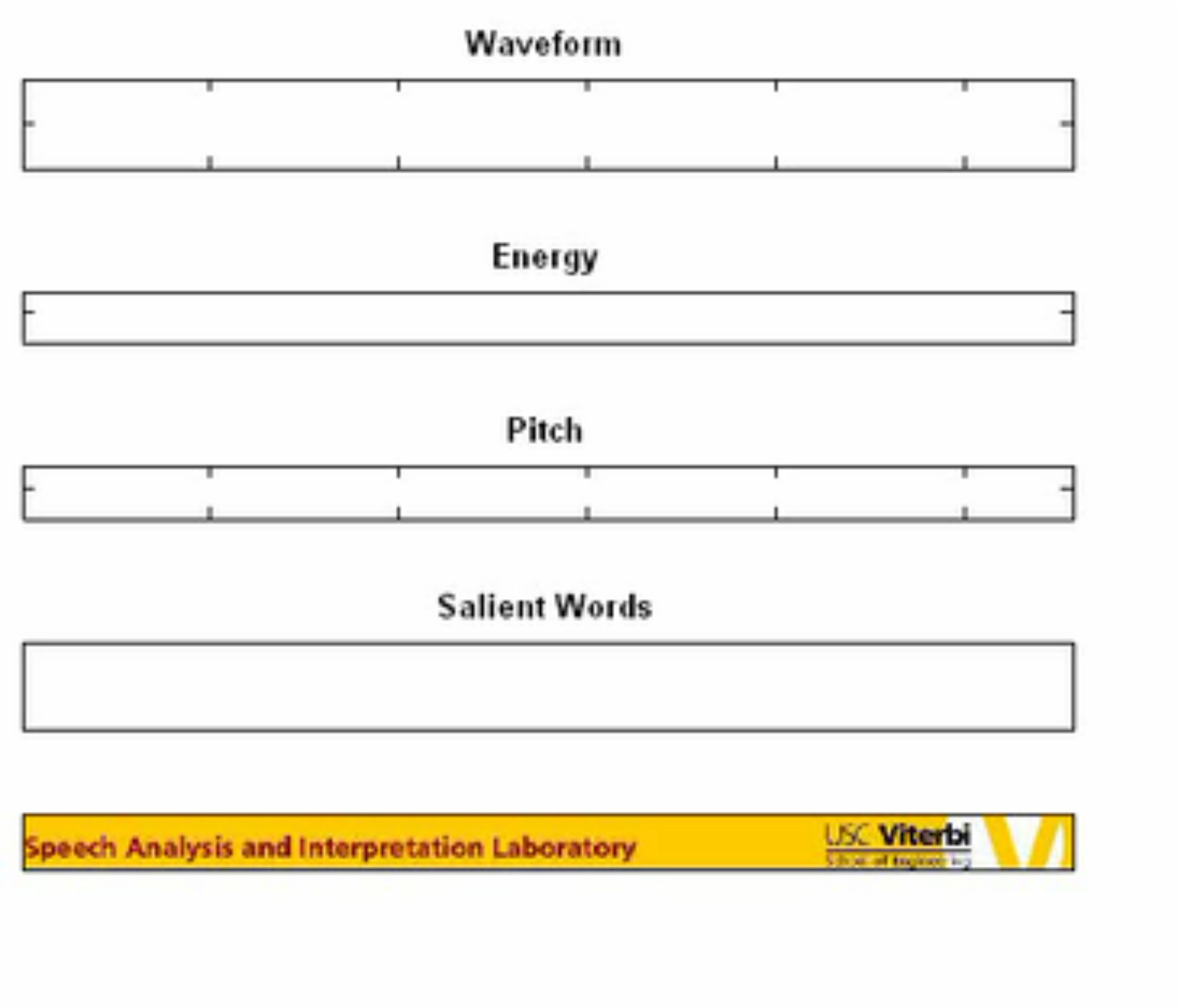
.. what data we need, how to process them, derive constructs for decision making

- **The phenomenon of interest: human data**
  - Behavior *Expression*, *Interaction* and *Judgment*

- **Purpose**
  - Scientific understanding, Technology development, Application design e.g., clinical translation

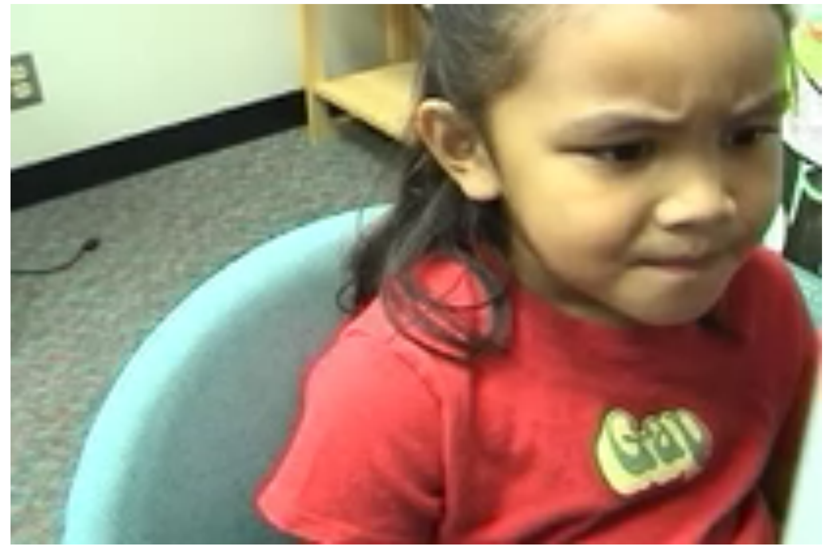
- **Use context**
  - Laboratory to Field studies to Real world environments
  - Contend with associated human and physical constraints

Customer care
Escalating frustration? (only customer side played)



Educational Game: "Cognitive state" Characterization

CONFIDENT vs. **UNCERTAIN**



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

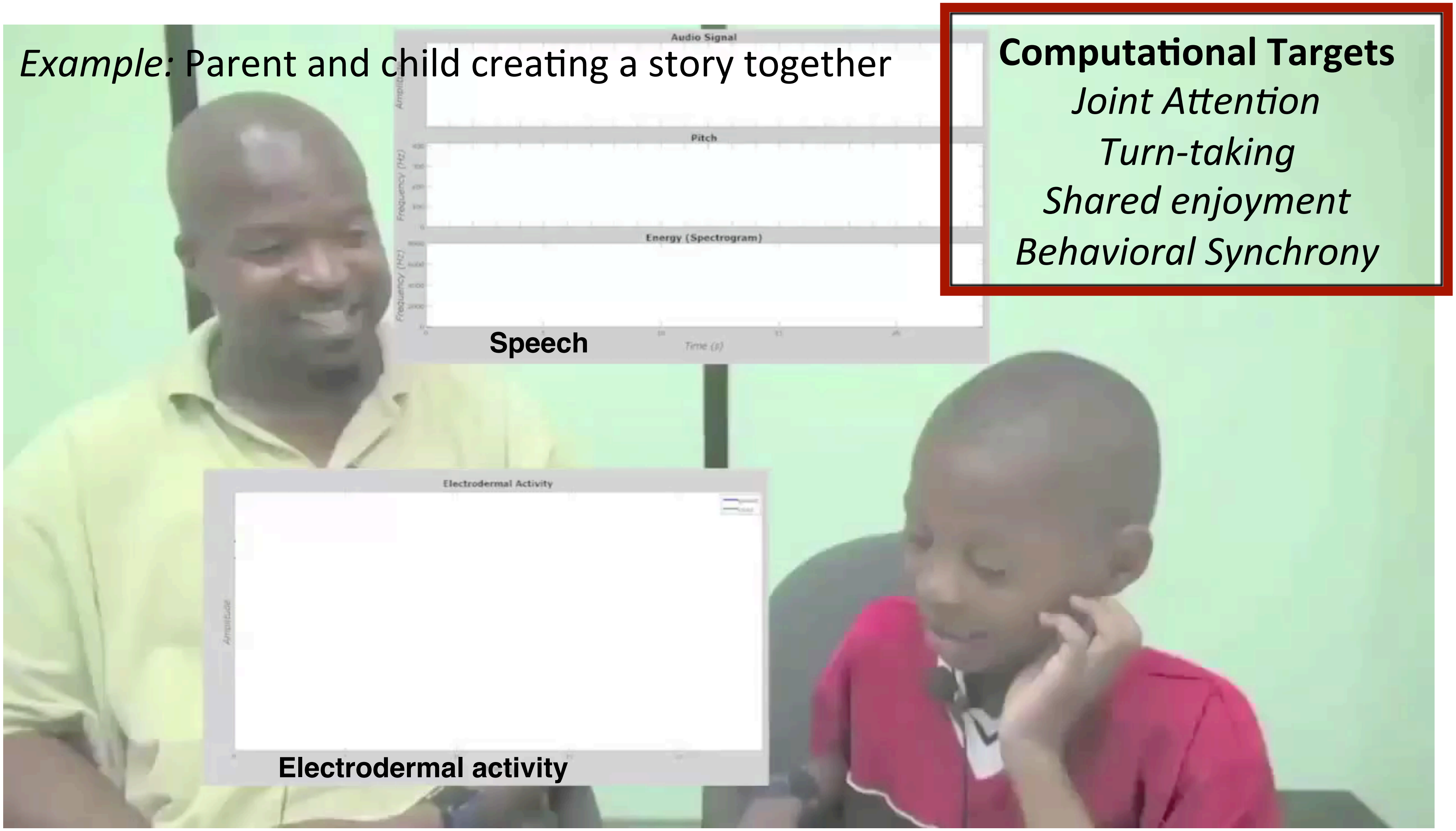
Distressed couple interactions: marital therapy

Characterizing affective dynamics, humor, blame patterns

CHRISTENSEN/CIRS TRAINING

Autism Spectrum Disorders

Technologies for Rich Understanding of Expressive Behavior and Interaction



Multimodal Behavior Signals

- Provide a window into internal state & processes

  Some overtly expressed and directly observable

  e.g., vocal and facial expressions, body posture

  Others, covert

  e.g., heart rate, electrodermal response, brain activity

- Implications for understanding
  - Human information encoding and decoding
  - "Mind-Body" relations
  - People's judgment of others behavior

MEASURING & QUANTIFYING HUMAN BEHAVIOR:
A CHALLENGING ENGINEERING PROBLEM

Operationally defining

Behavioral Signal Processing (BSP)

COMPUTATIONAL METHODS THAT MODEL HUMAN BEHAVIOR SIGNALS

- **Manifested In Both Overt And Covert Signals**
- **Processed And Used By Humans Explicitly Or Implicitly**
- **Facilitate Human Analysis And Decision Making**

OUTCOME OF BSP: "<u>BEHAVIORAL INFORMATICS</u>"

QUANTIFYING
HUMAN EXPRESSED BEHAVIOR
AND
HUMAN "FELT SENSE"

How is technology helping 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 signal processing with EKG, GSR, ..

SIGNAL PROCESSING AND MACHINE LEARNING ARE KEY ENABLERS

Example: A whole 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

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)**
Gender:	**Male**
Age:	**Adult**
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

How is technology helping 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 signal processing with EKG, GSR, ..

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

Ongoing Advances: Multifaceted

- **Sensing: From Smartrooms to Body area networks (instrumented people "in the wild")**

- **Rich speech/spoken language and video understanding**
  - ***who* said *what* to *whom, how* and *when & where***
- **Affective computing & Emotion modeling**
  - 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 BEHAVIORAL 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 environment

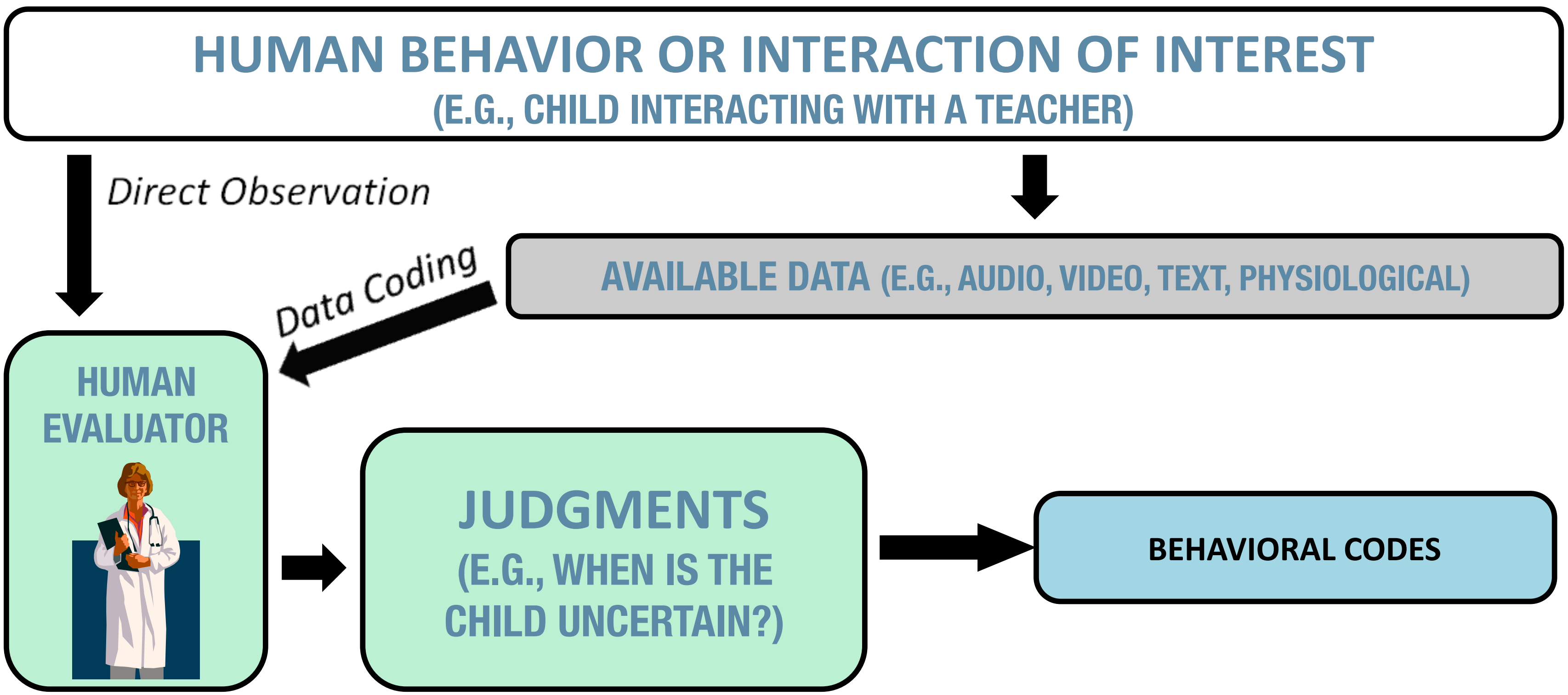
- **Analysis: deriving signal descriptors**
  - Deriving low level cues: **who, what, when, how, where, why**

- **Modeling: mapping behavioral constructs**
  - High level descriptions desired by domain experts
    - theory informed or to inform theory
  - Descriptive and predictive models using multimodal data
  - **Handle varying types of abstraction in data and descriptions**

- Heterogeneity and variability in how data are generated and used
- Uncertainty in observations (partial, noisy)
- Subjectivity in descriptions (especially of higher level behavior)

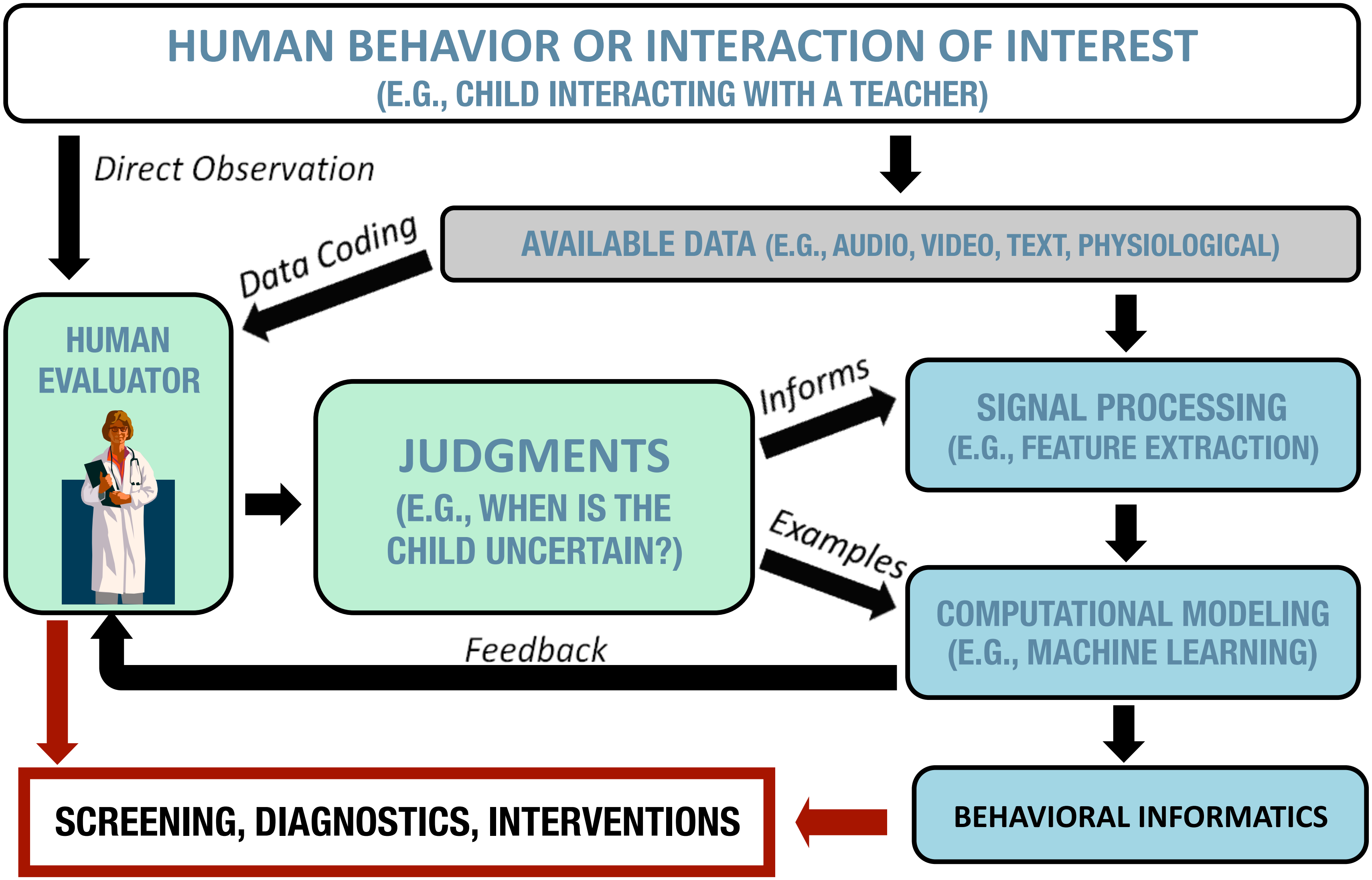
Behavior Coding: Humans in the loop

- Human assessments/judgments on human behavior



Behavior Coding: Humans in the loop

- Support-than supplant-human (expert) analyses



Behavioral signal processing: 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

TALK OUTLINE

Some behavioral informatics building blocks

- Focus on multimodal data processing
- Affective Computing as an example

Some Case Studies

- Dyadic interaction of distressed couples
    - Marital therapy
- Autism Spectrum Disorders
    - Quantifying social interaction and communication: Diagnostics, Outcomes
- Addiction
    - Understanding and evaluating psychotherapy

Multimodal data & processing techniques crucial for computational studies of behavior

Affective behavior computing as an example…

The Call center Corpus

Human-Computer Agent telephone interactions

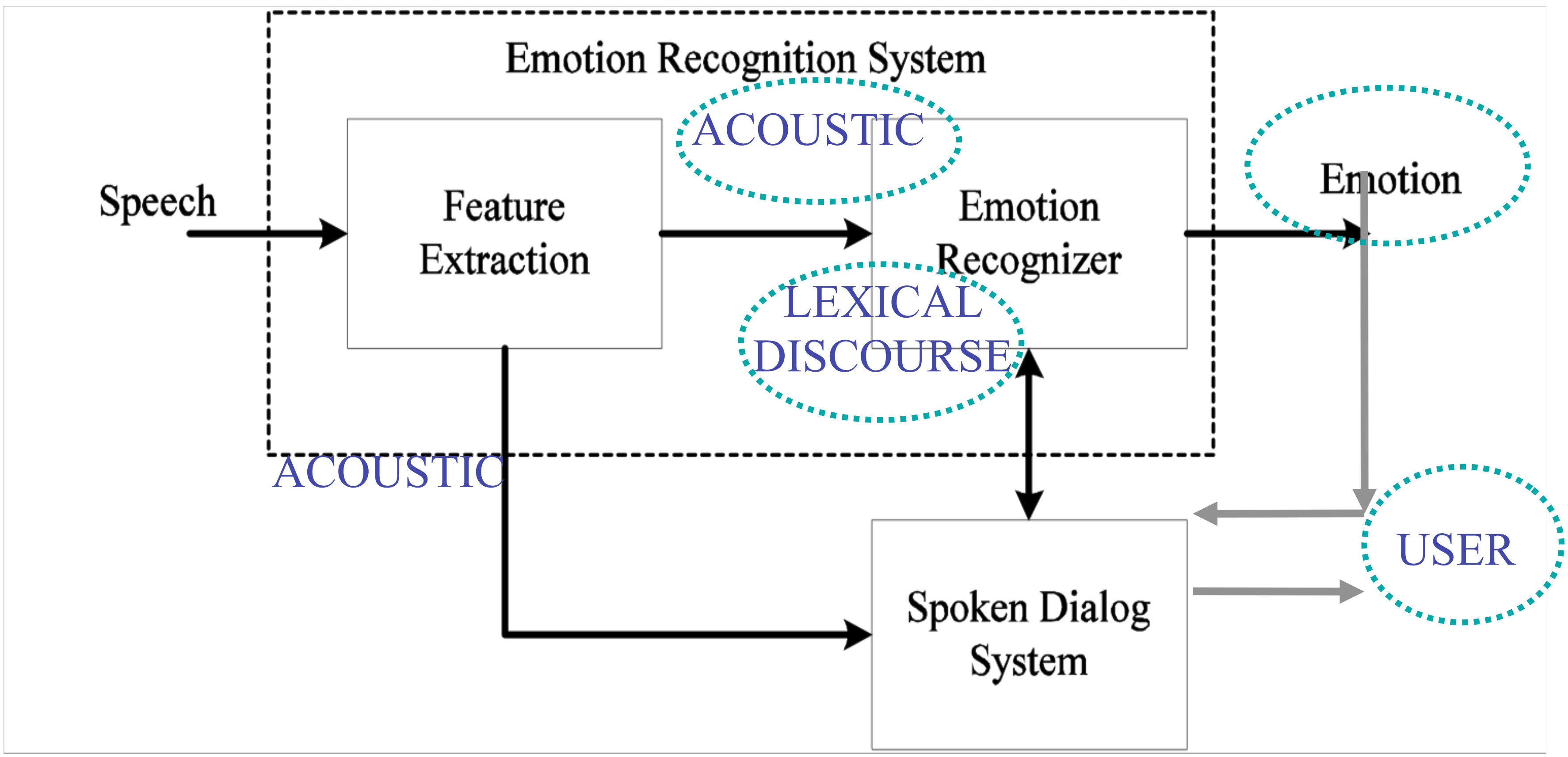
- Spoken dialog, emotions research
- Natural, spontaneous interactions; limited domain
- Categorical, dimensional ratings

Computing Emotions?

Expression *versus* Experience *versus* Judgment

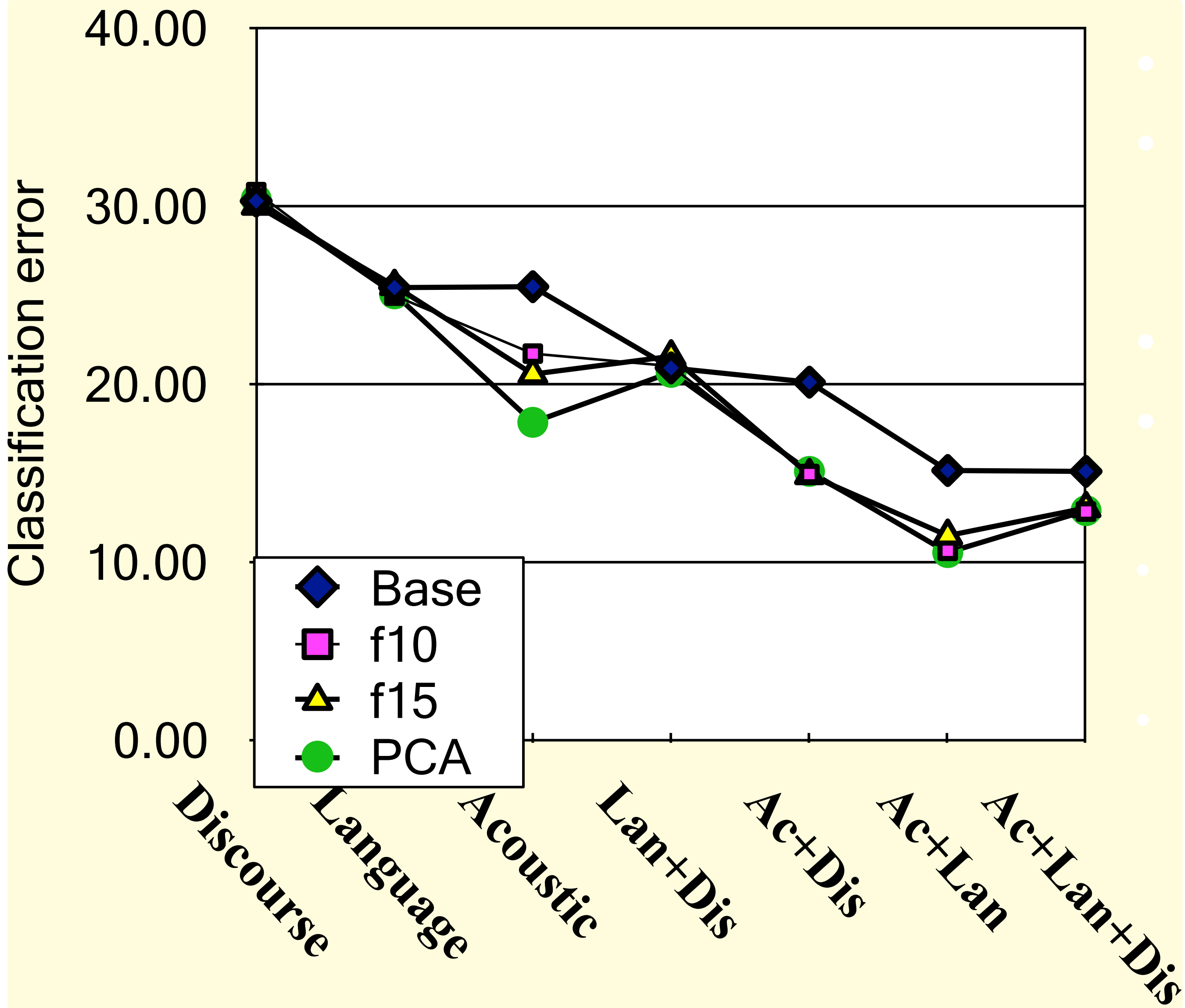
Representations for computation:

- Categorical (e.g.,happy, sad), Dimensional (arousal, valence, dominance)
- 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 PROCESSING, 13(2):293-302, 2005

Recognizing frustration



- **Call center data**
- **Classification method**
  - Linear discriminant classifier for each information modality
- **Modalities**
- Acoustic features
  - prosody, spectral features
- Language features
  - emotional salience of words
- Discourse
  - dialog acts

LEE, NARAYANAN & PIERACCINI. RECOGNITION OF NEGATIVE EMOTIONS FROM THE SPEECH SIGNAL. PROC. IEEE AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING, DECEMBER 2001.

The VAM Corpus

A multimodal corpus of talk show interactions
(Karlsruhe, USC)

- Computational modeling, different annotation perspectives
- Incidental: (Sort of) natural, human interaction based, spontaneous
- Categorical, dimensional ratings

VAM Corpus details

Freely available: http://emotion-research.net/download/vam

- Unscripted discussions between talk-show guests
  - German; 47 speakers (11 m/36 f)
  - 893 utterances, average duration: 3.0 s
  - Audio, Video, Faces
- More authentic emotions
  - Many negative emotions
  - Text-free, icon-based evaluation using Self Assessment Manikins
  - Categorical & Dimensional evaluation by 6 German and 17 non German evaluators

MICHAEL GRIMM, KRISTIAN KROSCHEL, AND SHRIKANTH NARAYANAN. THE VERA AM MITTAG GERMAN AUDIO-VISUAL EMOTIONAL SPEECH DATABASE. IN PROC. INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO (ICME), JUNE 2008.

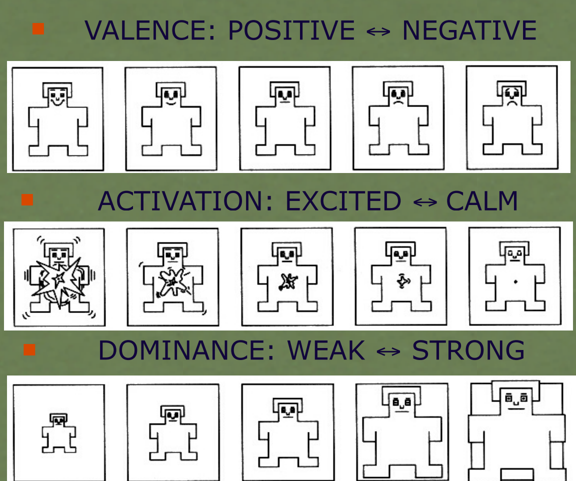
Emotion “Primitives”:
Valence, Activation, Dominance

CATEGORICAL EMOTIONS

EMOTIONS ARE DESCRIBED IN TERMS OF ... EMOTION CATEGORIES

... USUALLY 2 TO 6 CLASSES ARE DISTINGUISHED

... OFTEN EMOTIONS PORTRAYED BY ACTORS

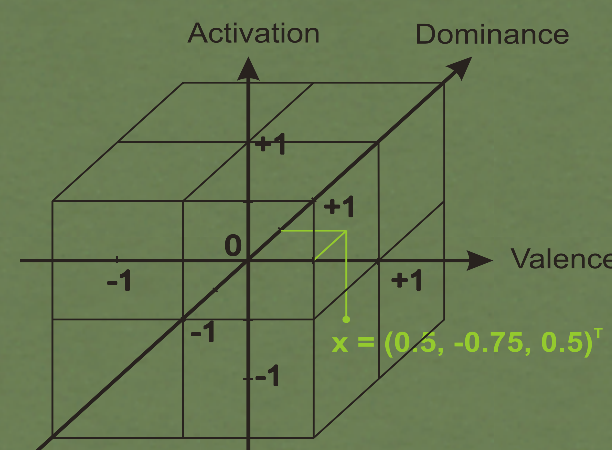


GRADIENT EMOTIONS

... EMOTIONS ARE DESCRIBED AS POINTS IN A 3D EMOTION SPACE

... EMOTIONS ARE ESTIMATED ON A CONTINUOUS-VALUED SCALE

... SPONTANEOUS EMOTIONS ARE USED



MICHAEL GRIMM, EMILY MOWER, KRISTIAN KROSCHEL, AND SHRIKANTH NARAYANAN. PRIMITIVES BASED ESTIMATION AND EVALUATION OF EMOTIONS IN SPEECH. SPEECH COMMUNICATION. 49: 787-800, NOVEMBER 2007.

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

- Multimodality
- Interaction dynamics

The USC IEMOCAP Corpus

A multimodal corpus of affective dyadic interactions

- Computational modeling, multimodal perspective
- Laboratory elicited (human interaction based), acted, spontaneous
- Categorical, dimensional ratings

Freely available: http://sail.usc.edu/iemocap

The USC IEMOCAP Database

Freely available: http://sail.usc.edu/iemocap

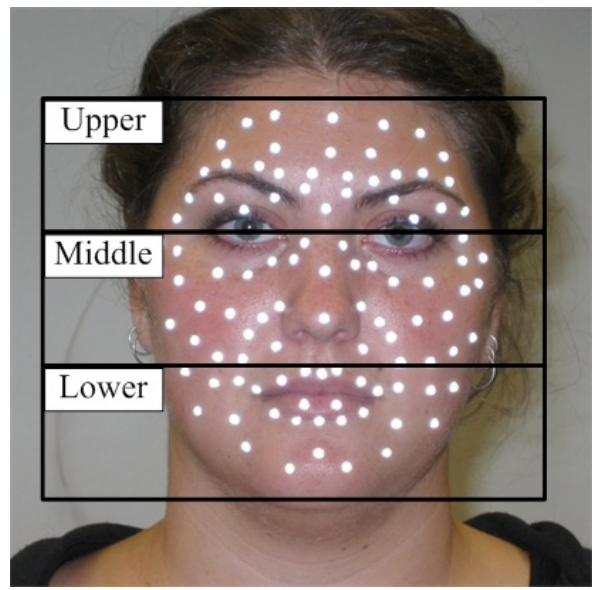
- Rich variety of emotions and multimodal manifestations in a dyadic interaction setting from actors
- Facial motion capture
  - 63 markers distributed on one actor’s face and hand
  - 3 Vicon Motion Capture Cameras
- Microphone speech
  - Shot gun directional microphones
- Video
  - 2 HD cameras directed at each actor
- Data collection settings
  - Spontaneous improvisations
  - Scripted improvisation based on plays

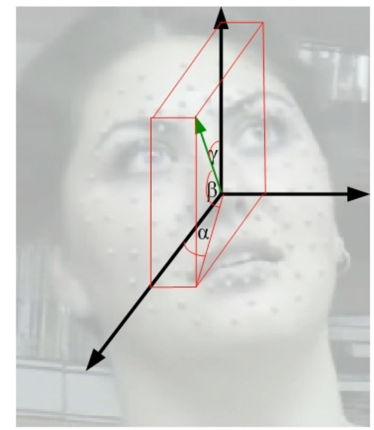
CARLOS BUSSO, MURTAZA BULUT, CHI-CHUN LEE, ABE KAZEMZADEH, EMILY MOWER, SAMUEL KIM, JEANNETTE CHANG, SUNGBOK LEE, AND SHRIKANTH NARAYANAN. IEMOCAP: INTERACTIVE EMOTIONAL DYADIC MOTION CAPTURE DATABASE. JOURNAL OF LANGUAGE RESOURCES AND EVALUATION. 42:335-359, NOVEMBER 2008.

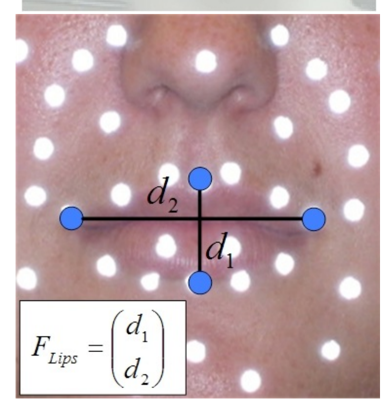
Modeling gesture/speech interrelation

VOCAL AND VISUAL FEATURES

- **Speech**
  - Prosodic features: Pitch, energy
  - MFCC coefficients (vocal tract)
- **Gestures**
  - 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

REDUNDANCY & COMPLEMENTARITY IN EMOTION ENCODING

- **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

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

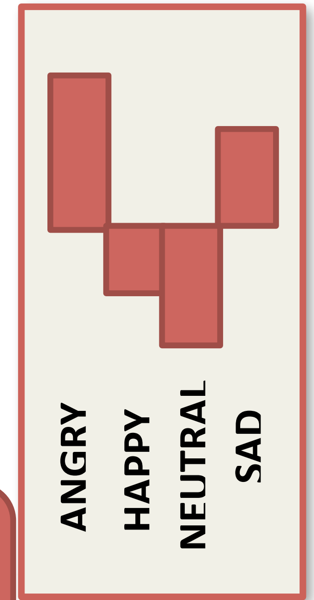
	Anger	Sadness	Happiness	Neutral
Anger	0.95	0.00	0.03	0.03
Sadness	0.00	0.79	0.03	0.18
Happiness	0.02	0.00	0.91	0.08
Neutral	0.01	0.05	0.02	0.92

BUSSO ET AL, ANALYSIS OF EMOTION RECOGNITION USING FACIAL EXPRESSIONS, SPEECH AND MULTIMODAL INFORMATION, ICMI, 2004

Profile-based Representations of Emotions

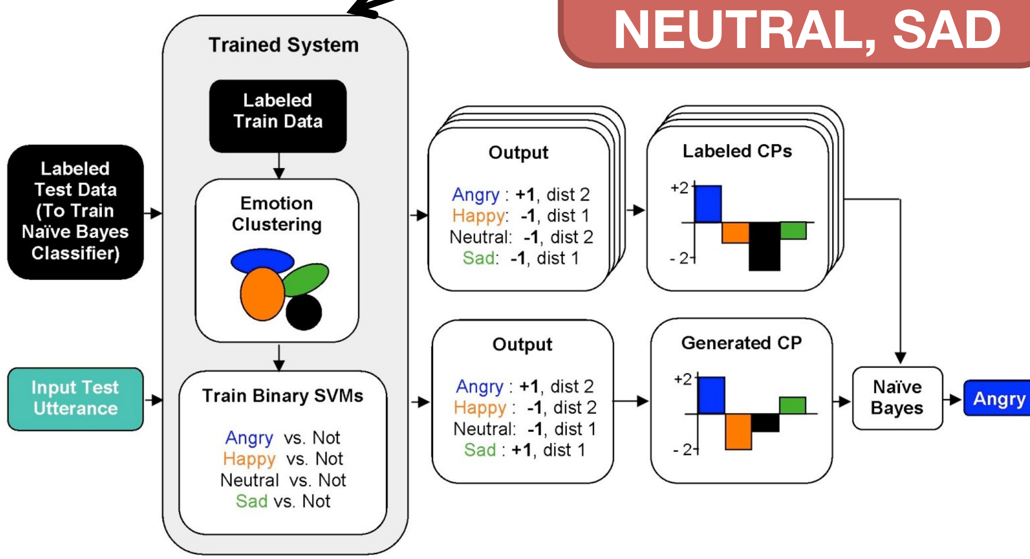
Characterizing Ambiguous Emotion Displays

Handling non-prototypical, blended emotions



TRAIN EPs ON: ANGRY, HAPPY, NEUTRAL, SAD

ANALYZE: ANGRY, HAPPY, NEUTRAL, SAD, <u>FRUSTRATED</u>

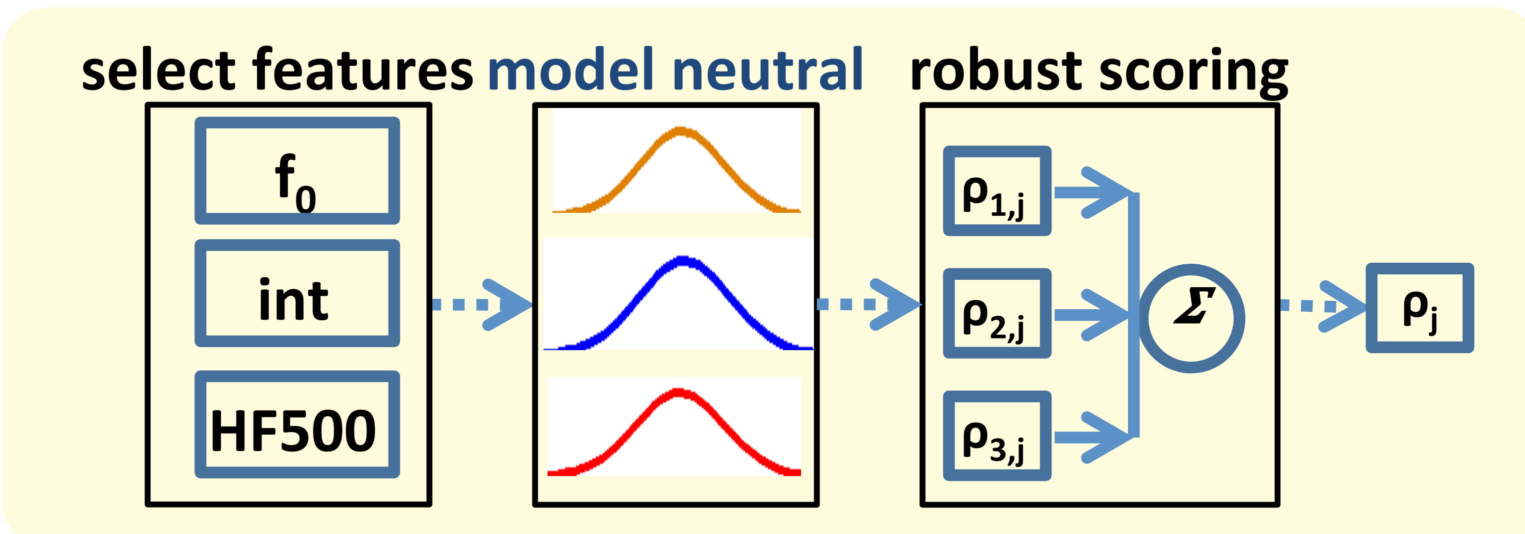


SUPERVISED OR UNSUPERVISED LEARNING VIA CLUSTERING OF THE EMOTION SPACE

EMILY MOWER, MAJA MATARIC AND SHRIKANTH NARAYANAN. A FRAMEWORK FOR AUTOMATIC HUMAN EMOTION CLASSIFICATION USING EMOTIONAL PROFILES. IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. 19(5): 1057-1070, 2011

Robust Arousal Estimation: A simple tool

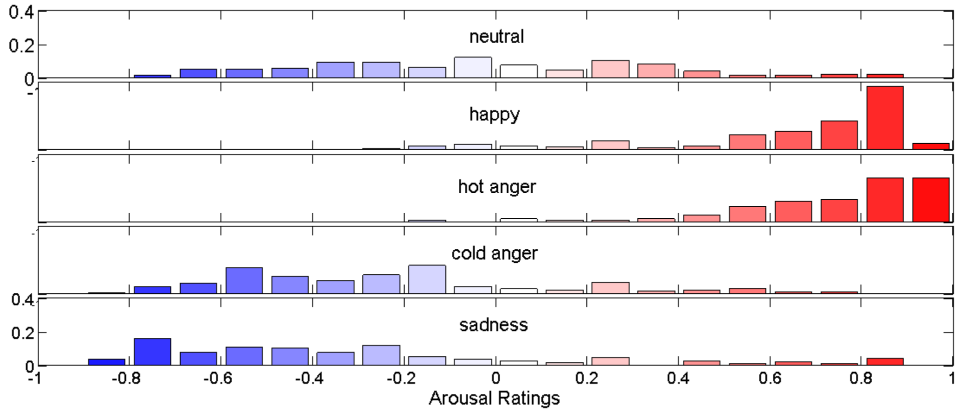
- Simple framework: compares favorably to cross-corpus supervised classification systems
- Tool generalizes and is simple: researchers can use to investigate behavioral hypotheses
  - 3 features: pitch, intensity, and spectral slope (HF500)
  - Chosen based on summative work of: Juslin and Scherer, The New Handbook of Methods in Nonverbal Behavior Research., 2005, ch. 3. Vocal Expression of Affect, pp. 65–135
- Largely unsupervised, only requires "neutral" labels from each speaker



DANIEL BONE, CHI-CHUN LEE AND SHRIKANTH S. NARAYANAN. ROBUST UNSUPERVISED AROUSAL RATING: A RULE-BASED FRAMEWORK WITH KNOWLEDGE-INSPIRED VOCAL FEATURES. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 5(1): 201-213, 2014.

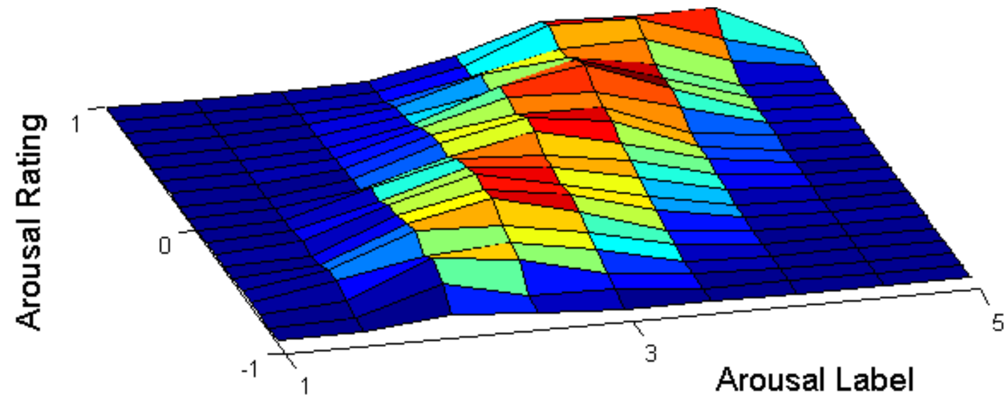
Robust Arousal Estimation

EMA Corpus

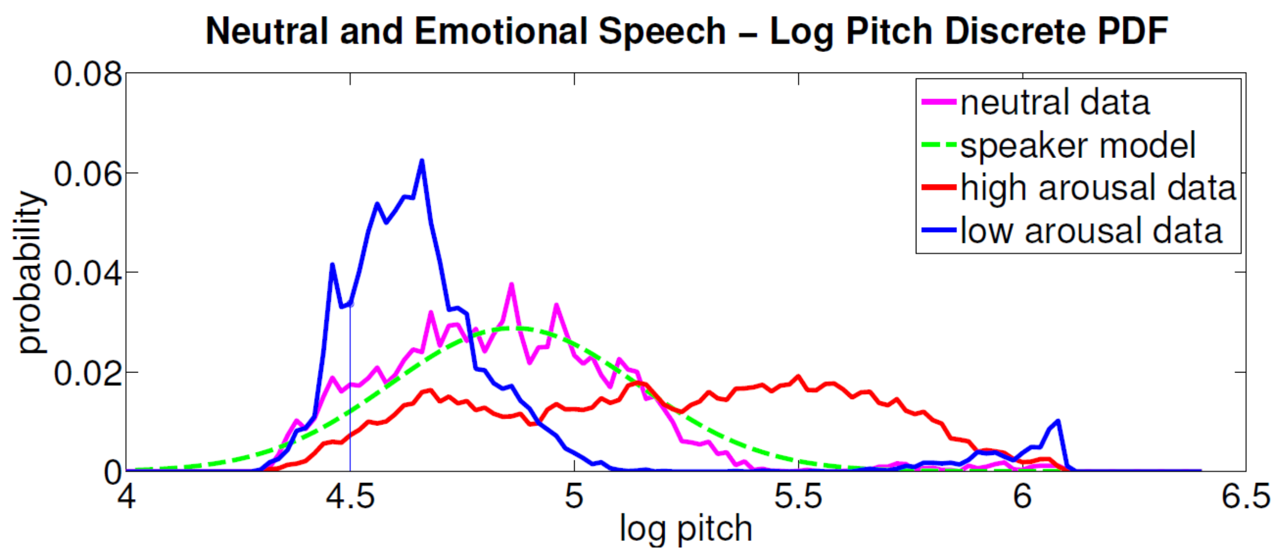


Automatic ratings cluster to hypothesized arousal levels for emotional categories

IEMOCAP Corpus



Automatic ratings correlate well with continuous manual arousal labels



Predictable shifts in certain features with arousal changes

Multimodal turn taking dynamics

Problem

- Incorporate "mutual influence" of interactants in the model

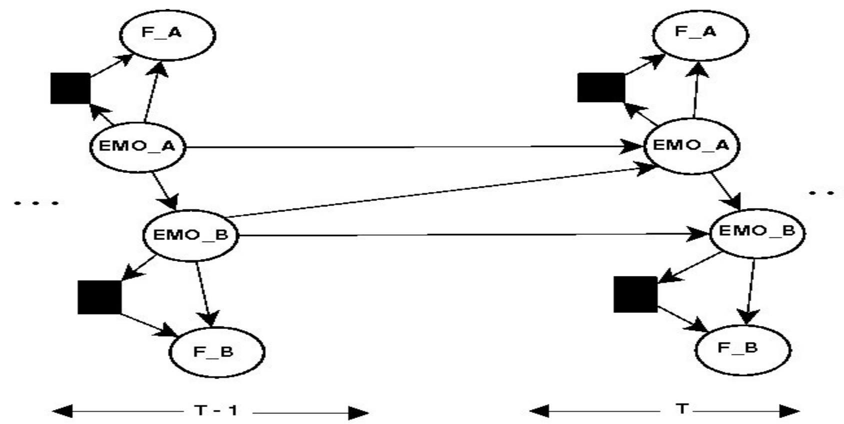
Approach

- 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

A. METALLINOU, M. WOLLMER, A. KATSAMANIS, F. EYBEN, B. SCHULLER, S. NARAYANAN. CONTEXT-SENSITIVE LEARNING FOR ENHANCED AUDIOVISUAL EMOTION CLASSIFICATION. IEEE TRANS. ON AFFECTIVE COMPUTING. 3: 184–198, 2012

The USC CreativeIT Corpus

A multimodal corpus of improv dyadic interactions

Freely available: http://sail.usc.edu/improv

- Computational modeling, multimodal (***full body mocap***) perspective
- Laboratory elicited (human interaction based), acted, spontaneous
- Categorical, dimensional ratings; ***continuous-time*** affect ratings

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. 2015

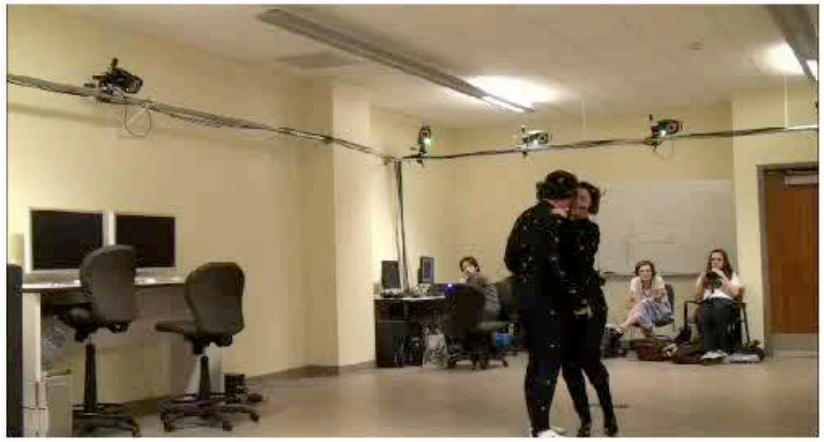
The USC Creative IT database

- Multimodal emotional database of theatrical improvisation
- Collaboration between engineering and theater
    - Active Analysis methodology **
    - Goal driven improvisations
- Rich variety of emotions and multimodal manifestations in an interaction setting
- Goals include:
    - Expressive body language and speech analysis
    - Emotion recognition
    - Study interaction dynamics
    - Animation of affective full-body virtual agents
    - Study actor's creativity/quality of performance

A. METALLINOU, C.-C. LEE, C. BUSSO, S. CARNICKE, AND S. NARAYANAN, "THE USC CREATIVEIT DATABASE: A MULTIMODAL DATABASE OF THEATRICAL IMPROVISATION," MULTIMODAL CORPORA, LREC, 2010
**S. M. CARNICKE, STANISLAVSKY IN FOCUS: AN ACTING MASTER FOR THE TWENTY-FIRST CENTURY, 1998

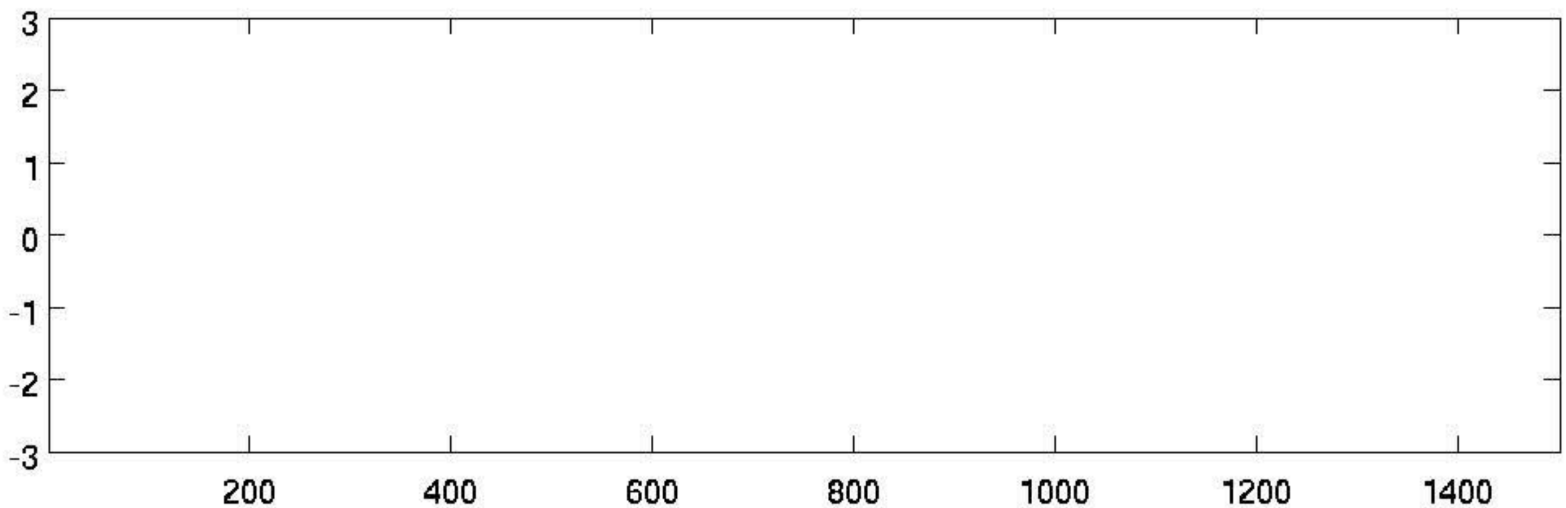
The USC Creative IT database

Freely available: http://sail.usc.edu/improv



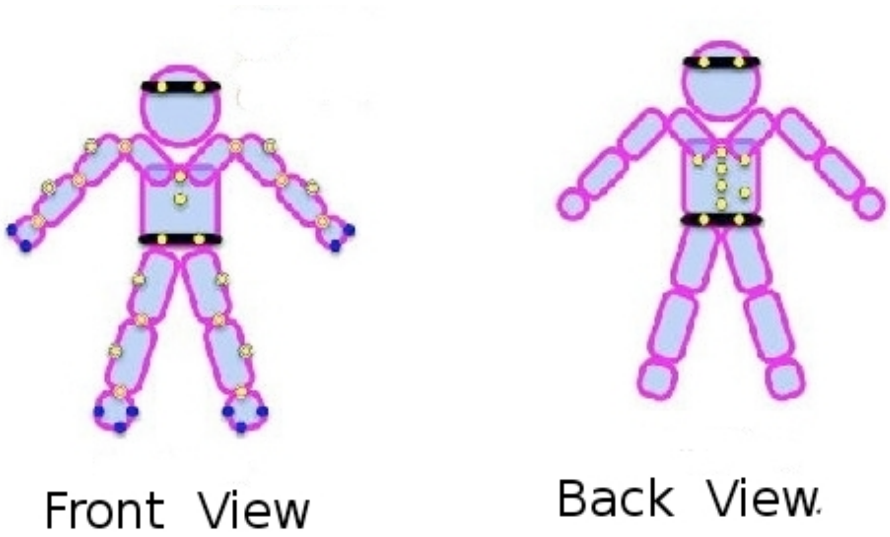
Continuous rating by three different annotators

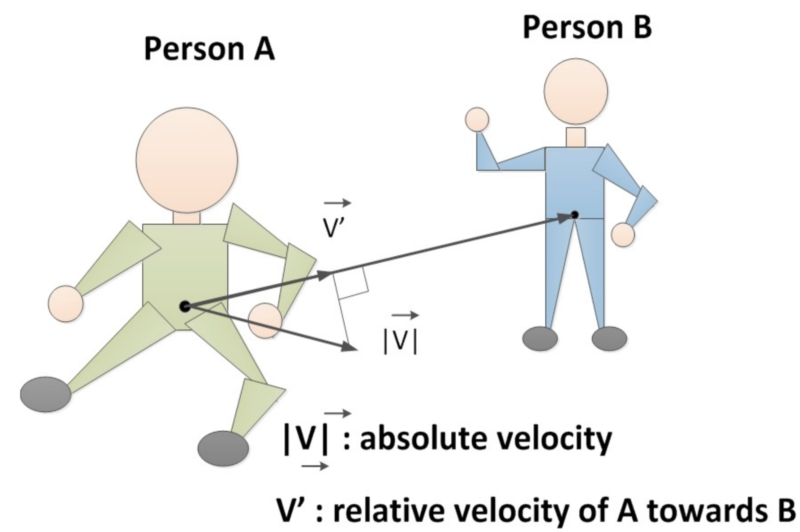
Activation of Male Actor

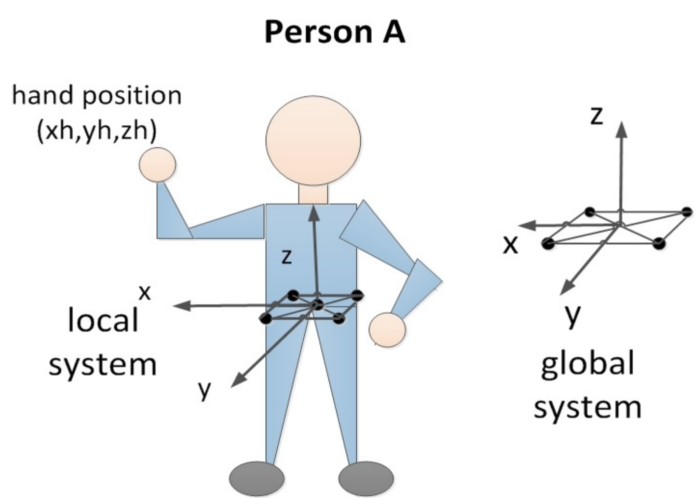


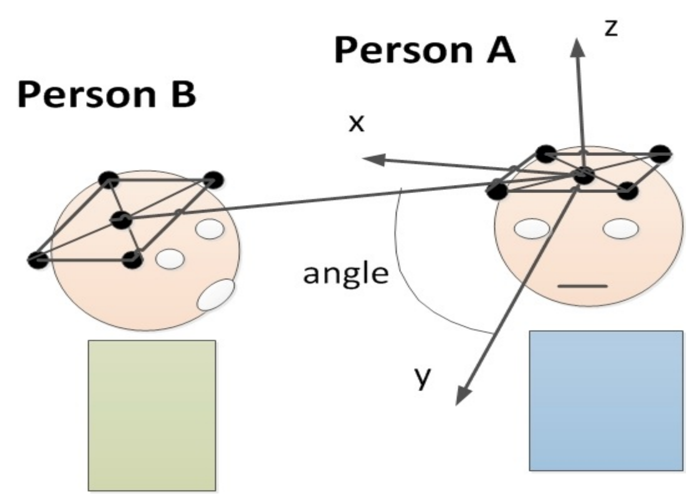
ANGELIKI METALLINOU AND S. NARAYANAN, ANNOTATION AND PROCESSING OF CONTINUOUS EMOTIONAL ATTRIBUTES: CHALLENGES AND OPPORTUNITIES, IN: 2ND INTERNATIONAL WORKSHOP ON EMOTION REPRESENTATION, ANALYSIS AND SYNTHESIS IN CONTINUOUS TIME AND SPACE (EMOSPACE), 2013

Body Language Feature Extraction

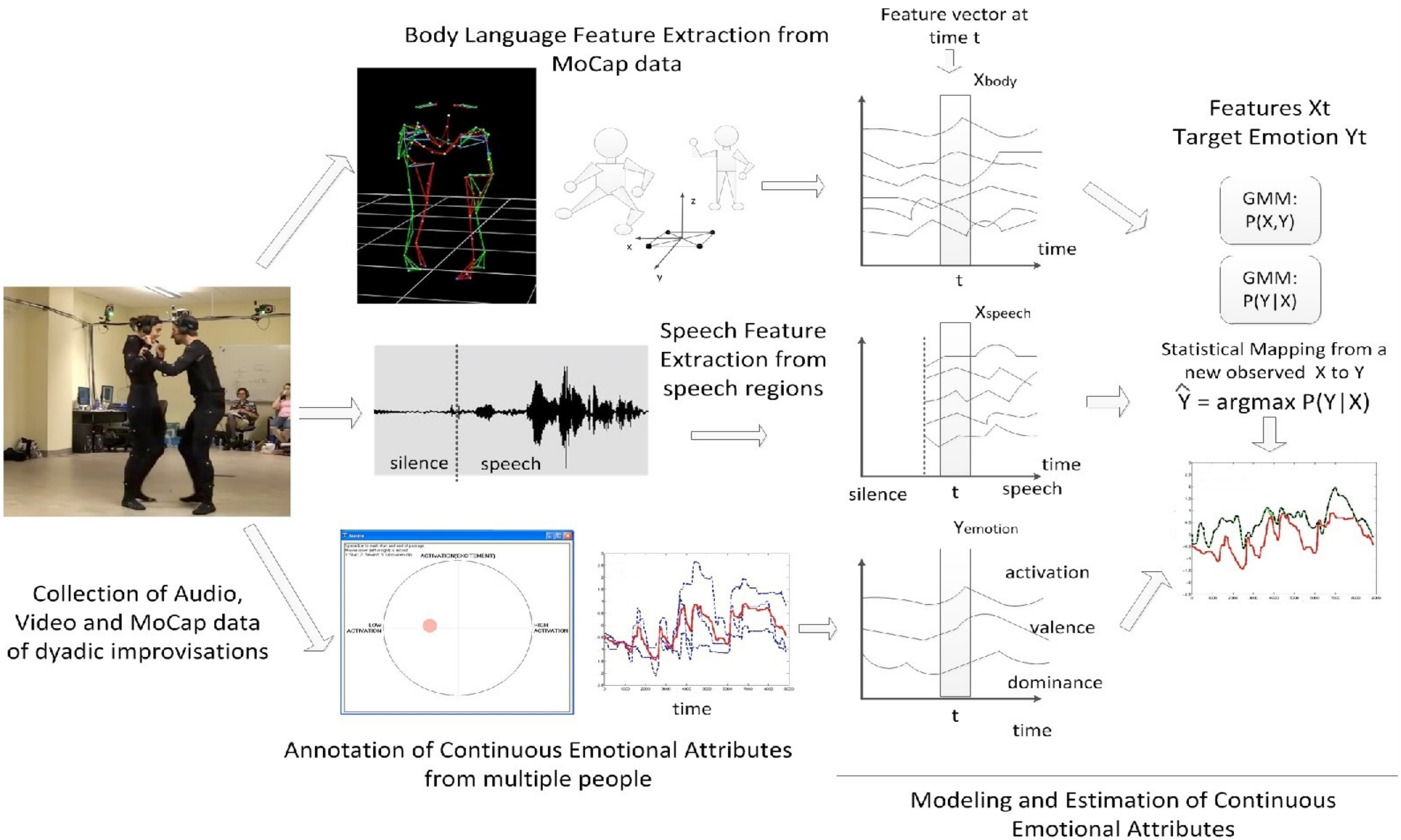








Tracking emotions from speech & body language



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, 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, 2014.

Modeling of Body Language Behavior from Multimodal Interaction Cues

- Focus on interactions where **friendly** versus **conflictive** stances are taken
- Expression of private internal state of attitude in multimodal cues

ZHAOJUN YANG, ANGELIKI METALLINOU, SHRIKANTH NARAYANAN, TOWARD BODY LANGUAGE GENERATION IN DYADIC INTERACTION SETTINGS FROM INTERLOCUTOR MULTIMODAL CUES, IN: PROCEEDINGS OF ICASSP, 2013

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, 2014.

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, 2013

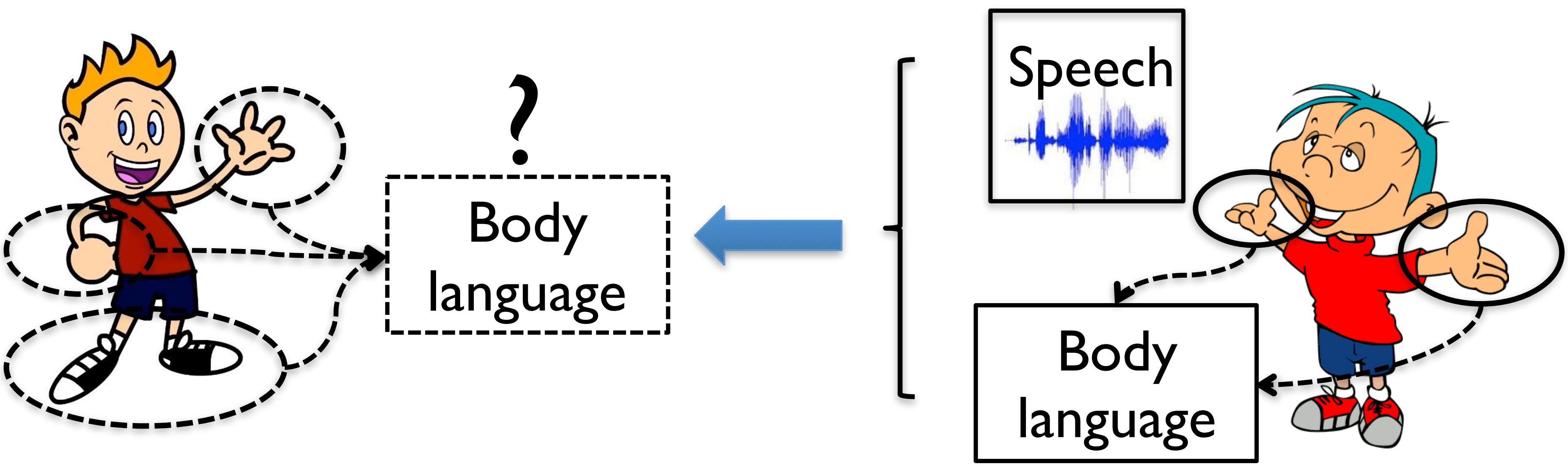
Predictive modeling

- Uncover the coordination patterns of dyad's behavior

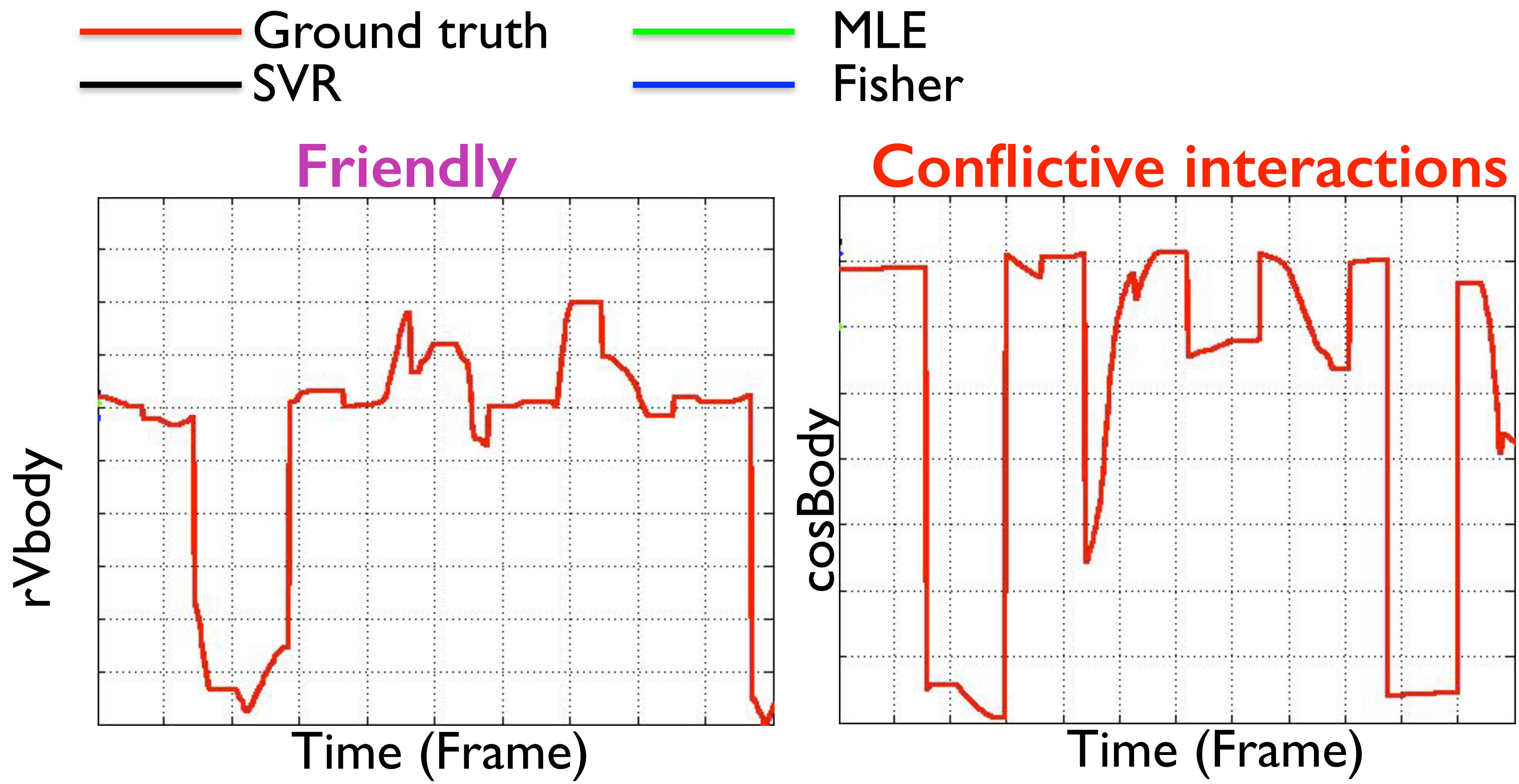
Friendly interactions

Conflictive interactions

- Computationally model the dyadic coordination



Predicted Body Language Trajectories



Fisher Kernel Approach: Good trend tracking & Estimation of values

higher predictability of body language orientation
in the friendly case →
increased adaptation to interlocutor behavior

Increased behavior coordination in positive affect
(also seen in vocal entrainment)

TALK OUTLINE

Some behavioral informatics building blocks

- Focus on multimodal data processing
- Affective Computing as an example

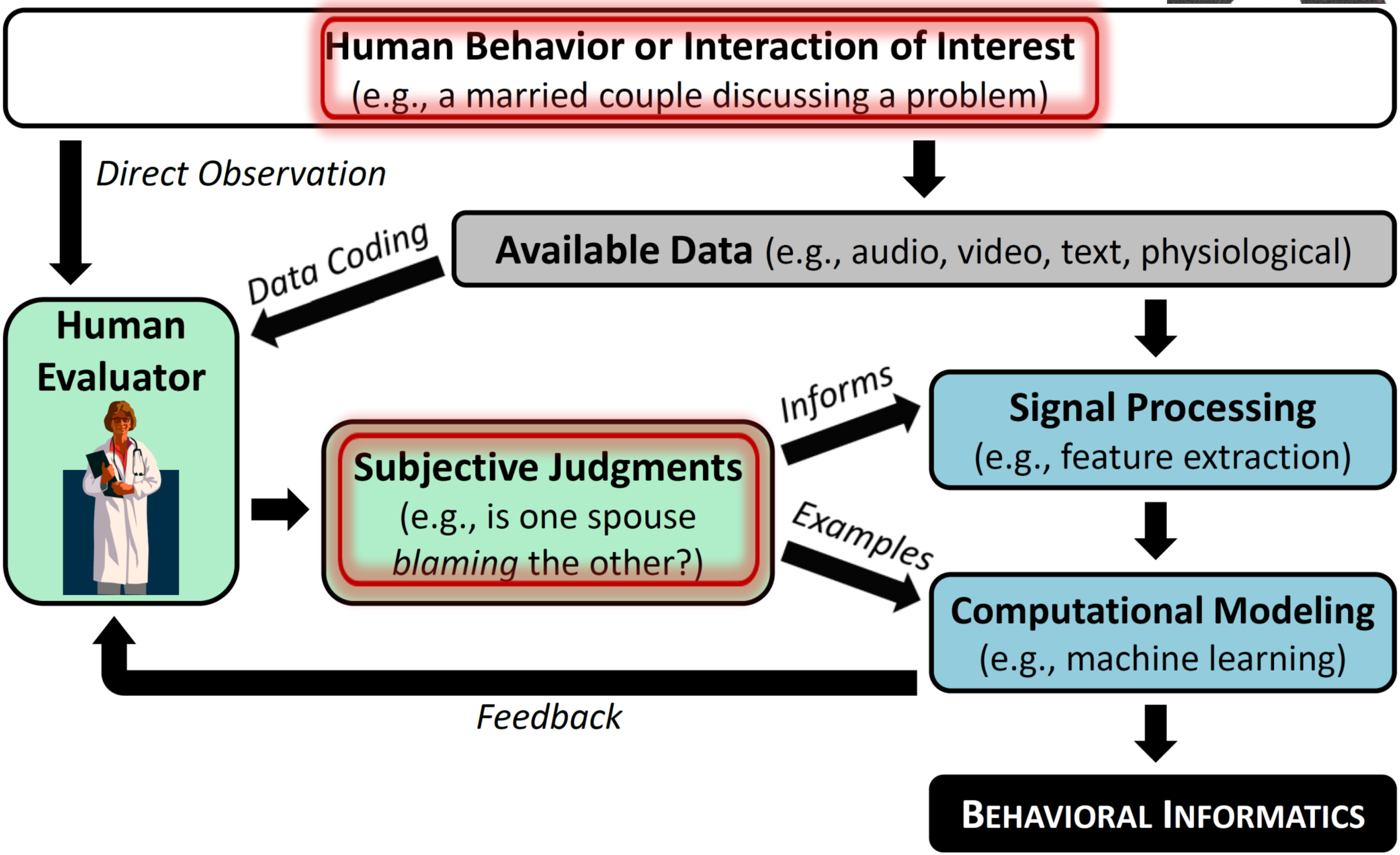
Some Case Studies

✓ Dyadic interaction of distressed couples

- Marital therapy

- Autism Spectrum Disorders
  - Quantifying social interaction and communication: Diagnostics, Outcomes
- Addiction
  - Understanding and evaluating psychotherapy

BSP for Couples Therapy Research



Couple therapy

Characterizing affective dynamics, blame patterns

Corpus

- **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
  - Video angle, microphone placement, and background noise varied

Manual 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

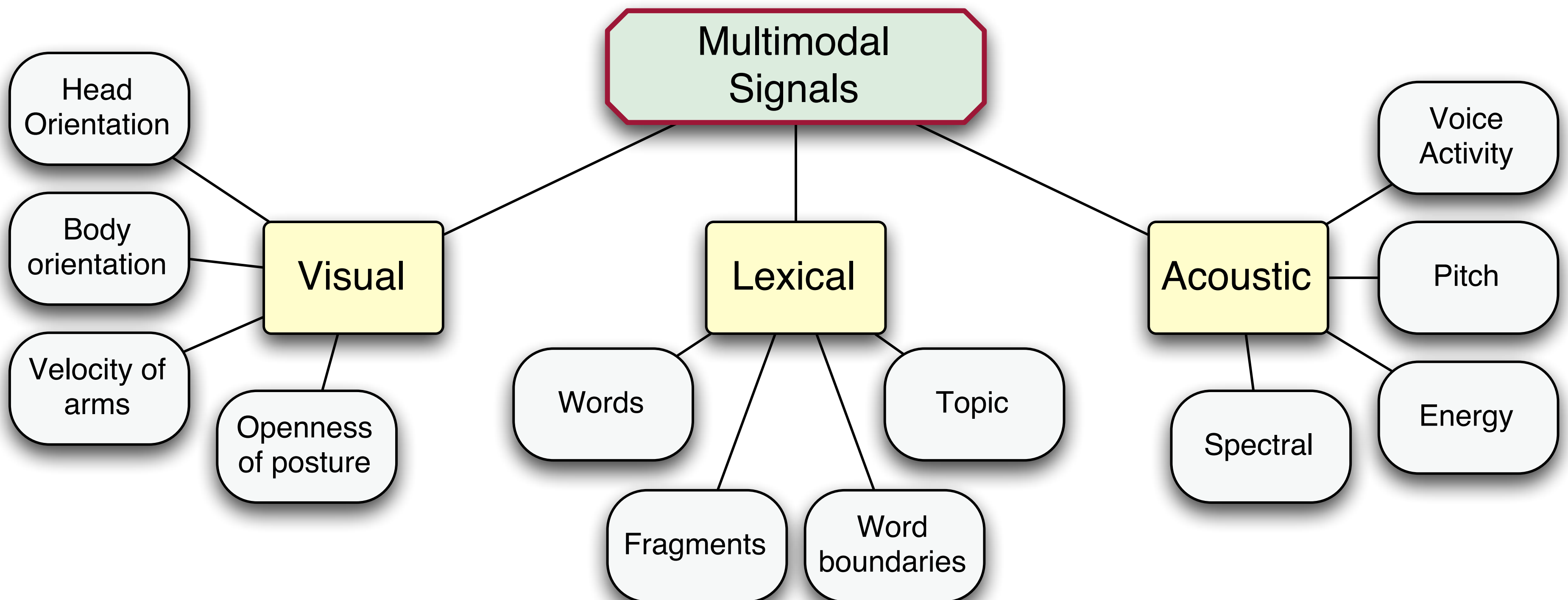
EXAMPLE CODING GOAL:
IS THE HUSBAND SHOWING ACCEPTANCE?" (SCALE 1-9)

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	Agreement
	acc	*bla*	*pos*	*neg*	*sad*	Correlation	
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

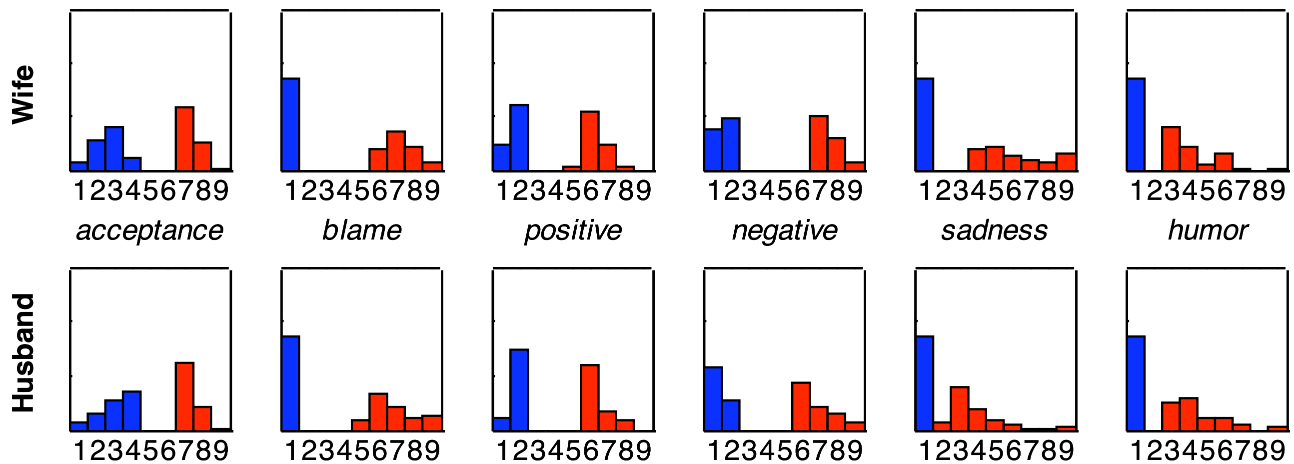
Automatic Behavior Coding:
Estimate behavioral codes from data



Focus on extreme cases of session-level judgments

Sample codes:

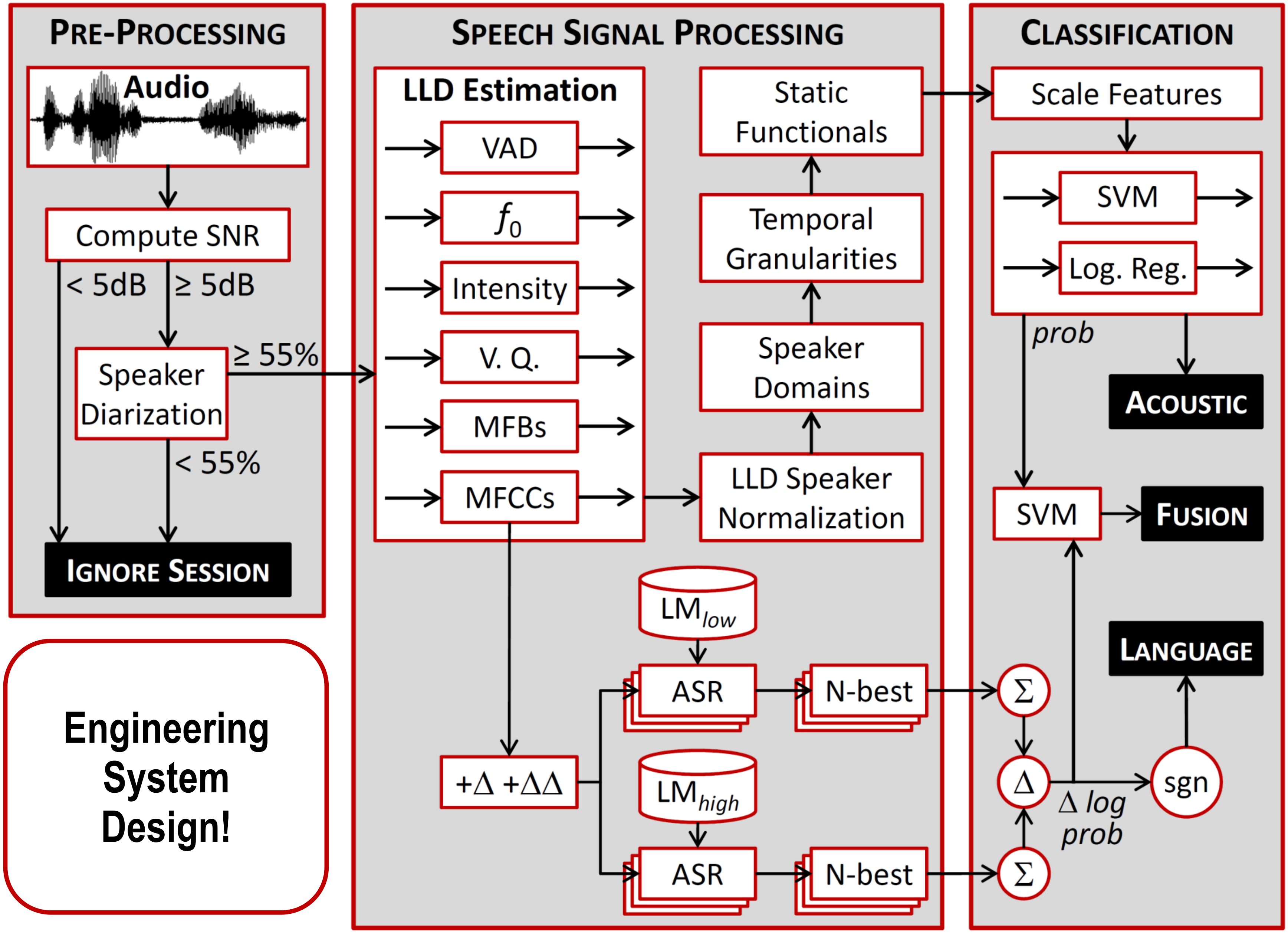
acceptance, blame, positive affect, negative affect, sadness, humor



M. Black, et al "Automatic classification of married couples' behavior using audio features" - Interspeech 2010

Matthew P Black, Athanasios Katsamanis, Brian R Baucom, Chi-Chun Lee, Adam C Lammert, Andrew Christensen, Panayiotis G Georgiou, Shrikanth S Narayanan. Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. Speech Communication. 55(1):1-21, 2013.

Methodology Pipeline

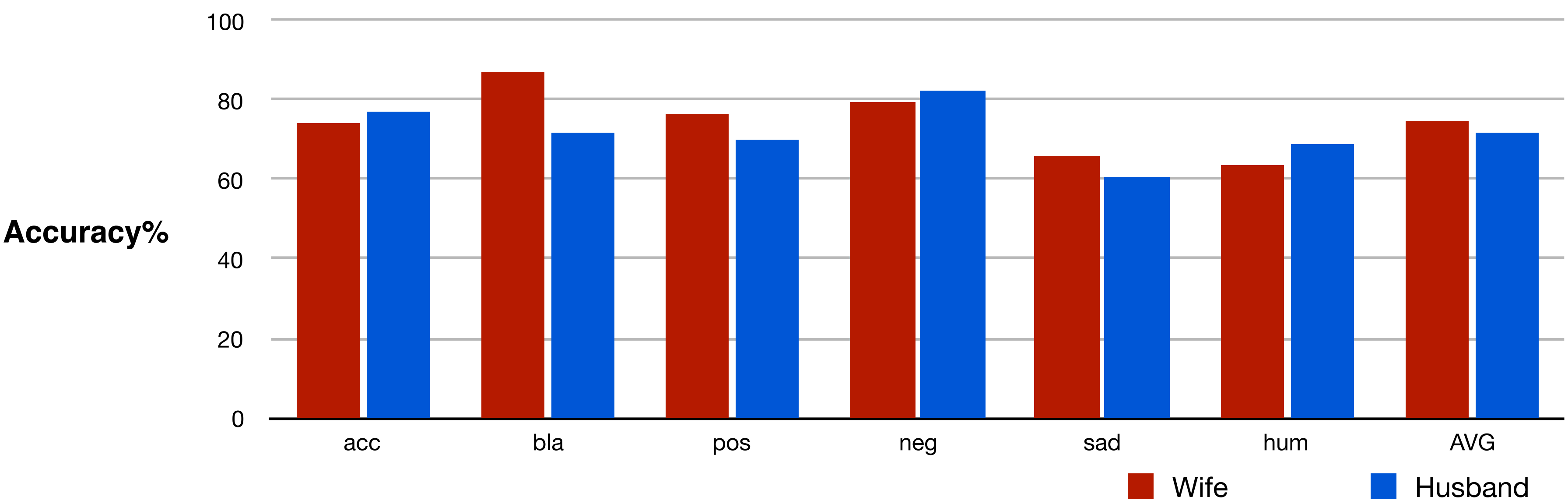


Engineering System Design!

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

(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

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"

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

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

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

Some technical challenges & 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**

Behavior Collection Space:
A new multichannel multimodal database

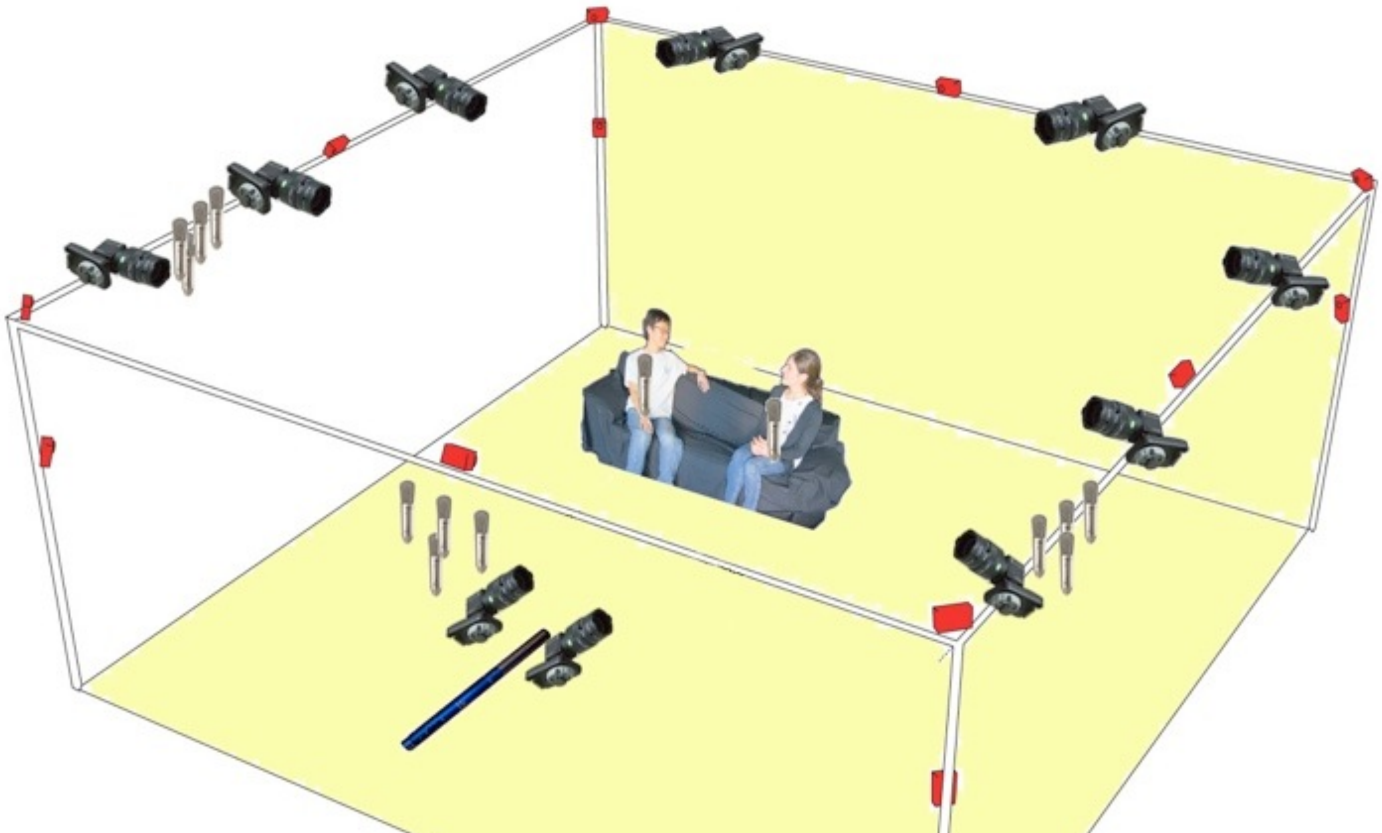
Audio:

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

Video:

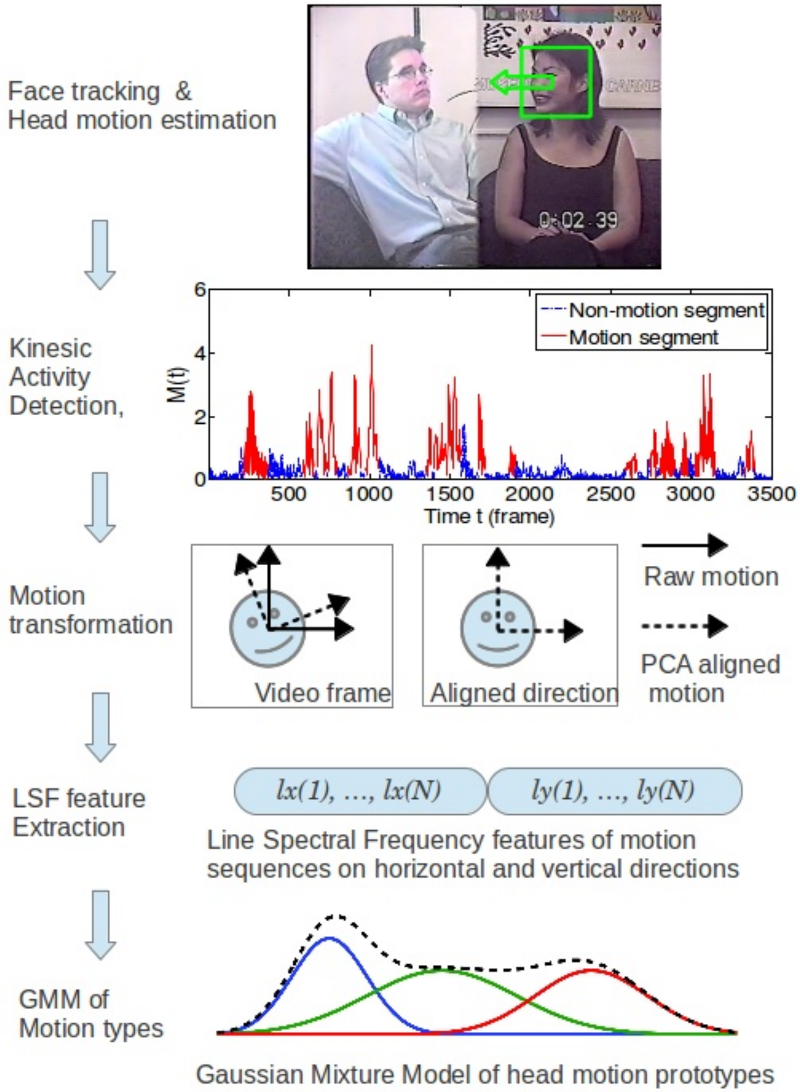
- 10 HD cameras (PointGrey Flea 2)
- Motion capture: 12 ViconQ Sensors

Accurate synchronization

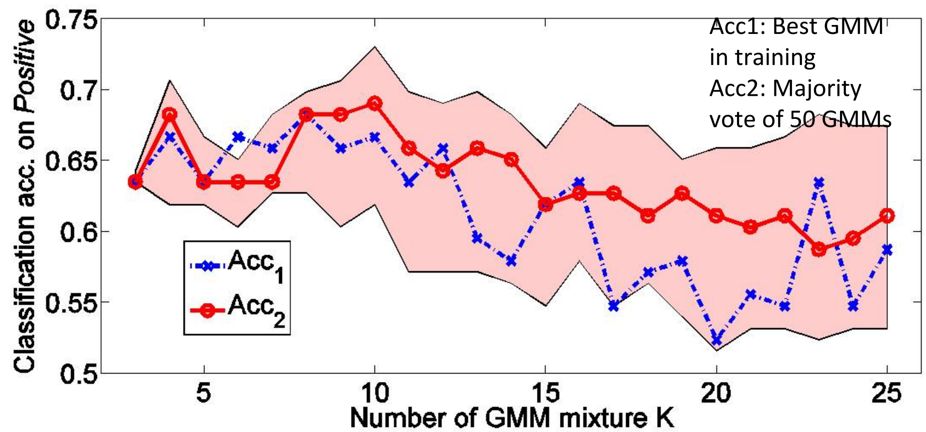


V. ROZGIĆ, B. XIAO, A. KATSAMANIS, B. BAUCOM, P. G. GEORGIOU, AND S. NARAYANAN, “A NEW MULTICHANNEL MULTIMODAL DYADIC INTERACTION DATABASE” INTERSPEECH 2010

Head motion modeling for behavior analysis



- **Head motion**
  - Important nonverbal behavior cues
  - Nods & shakes are common
- **Data driven modeling**
  - Optical flow of head motion
  - Motion segmentation
  - LSF representation
  - GMM clustering
  - Predict expert annotated behavior codes
  - Binary classification: ensemble of GMMs



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 (ICME), 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

Some technical challenges & 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

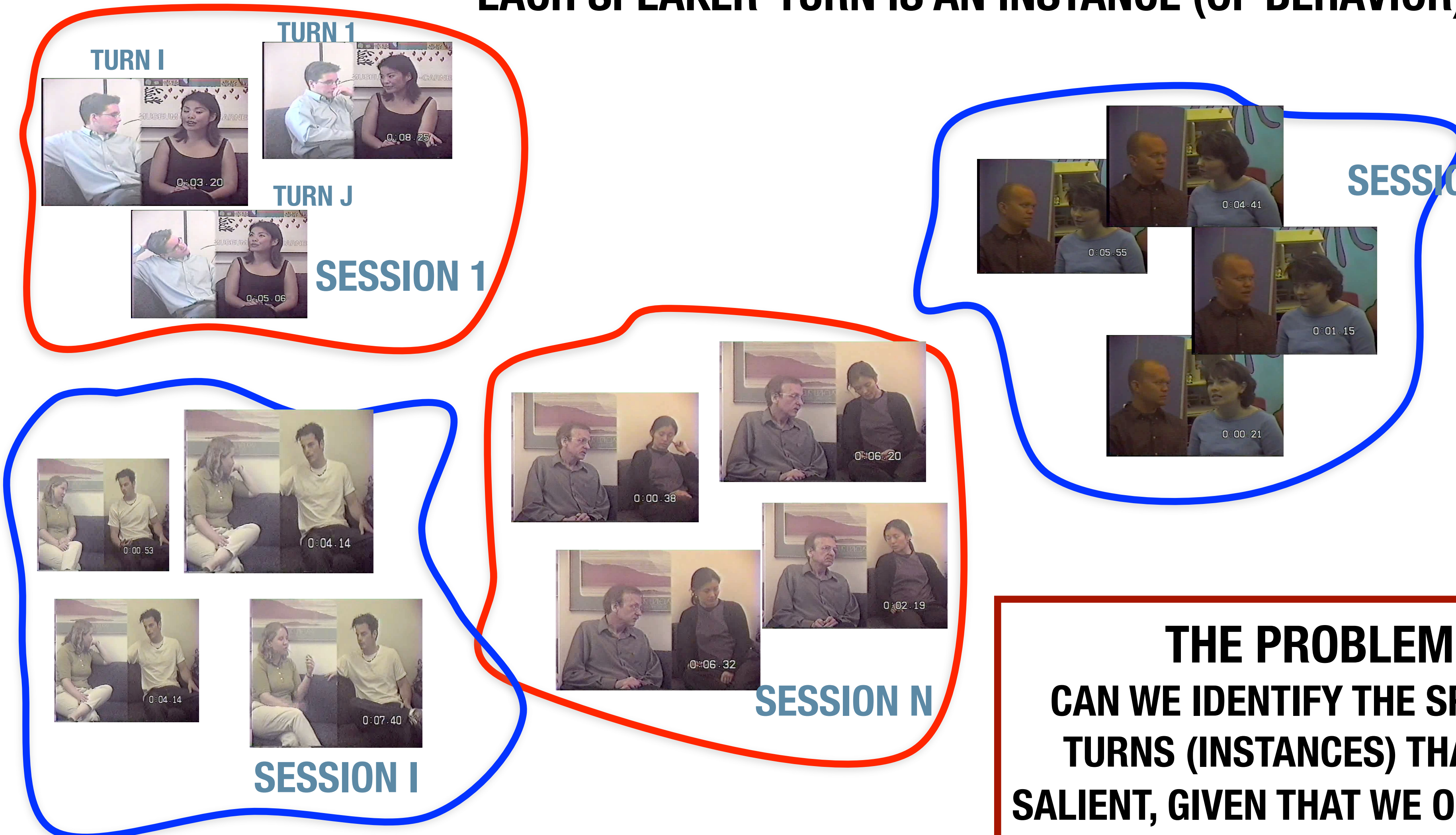
➡ **Models of entrainment**

- 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)

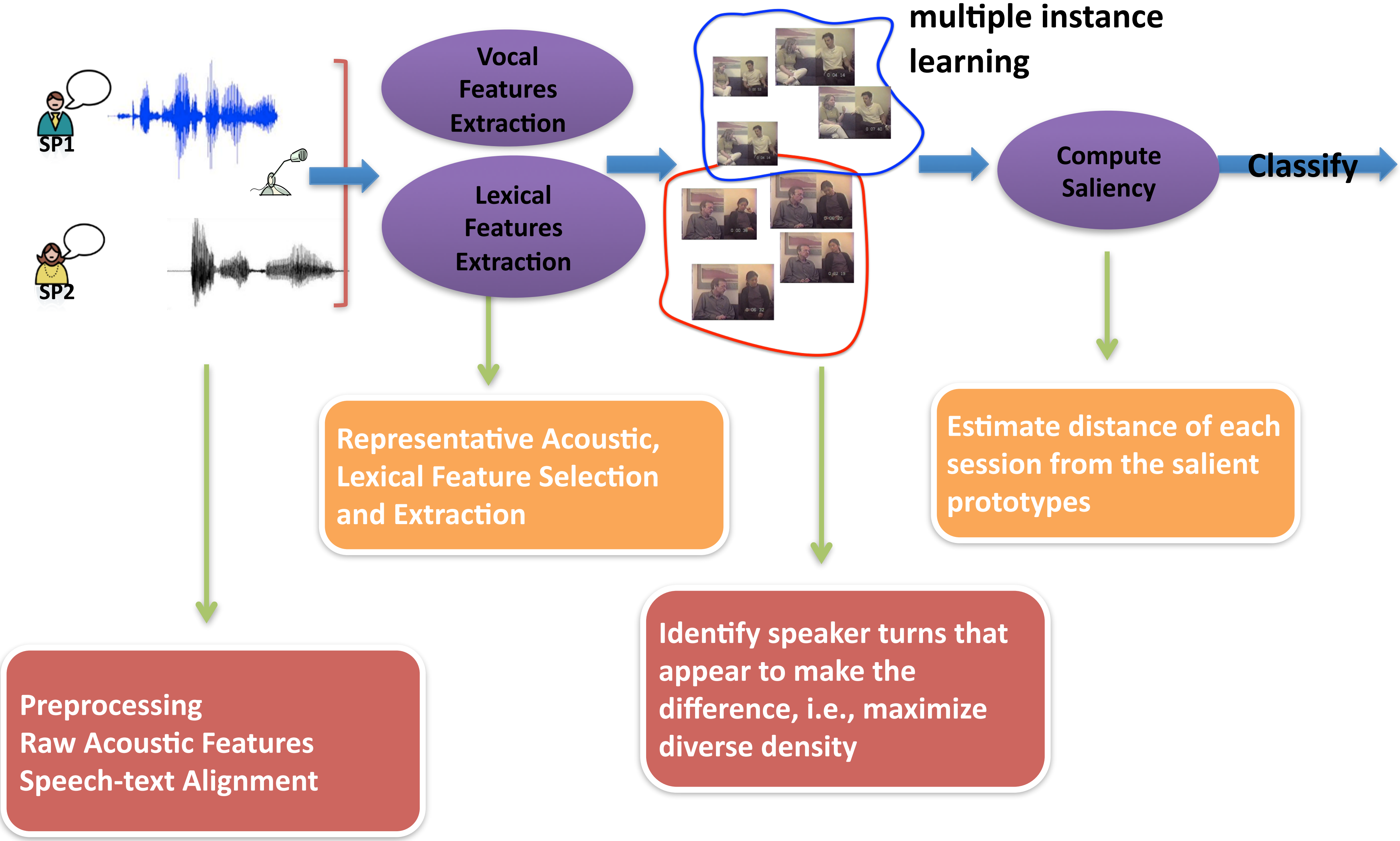


RED SESSIONS: NON-ACCEPTING SPOUSE
BLUE SESSIONS: ACCEPTING SPOUSE

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

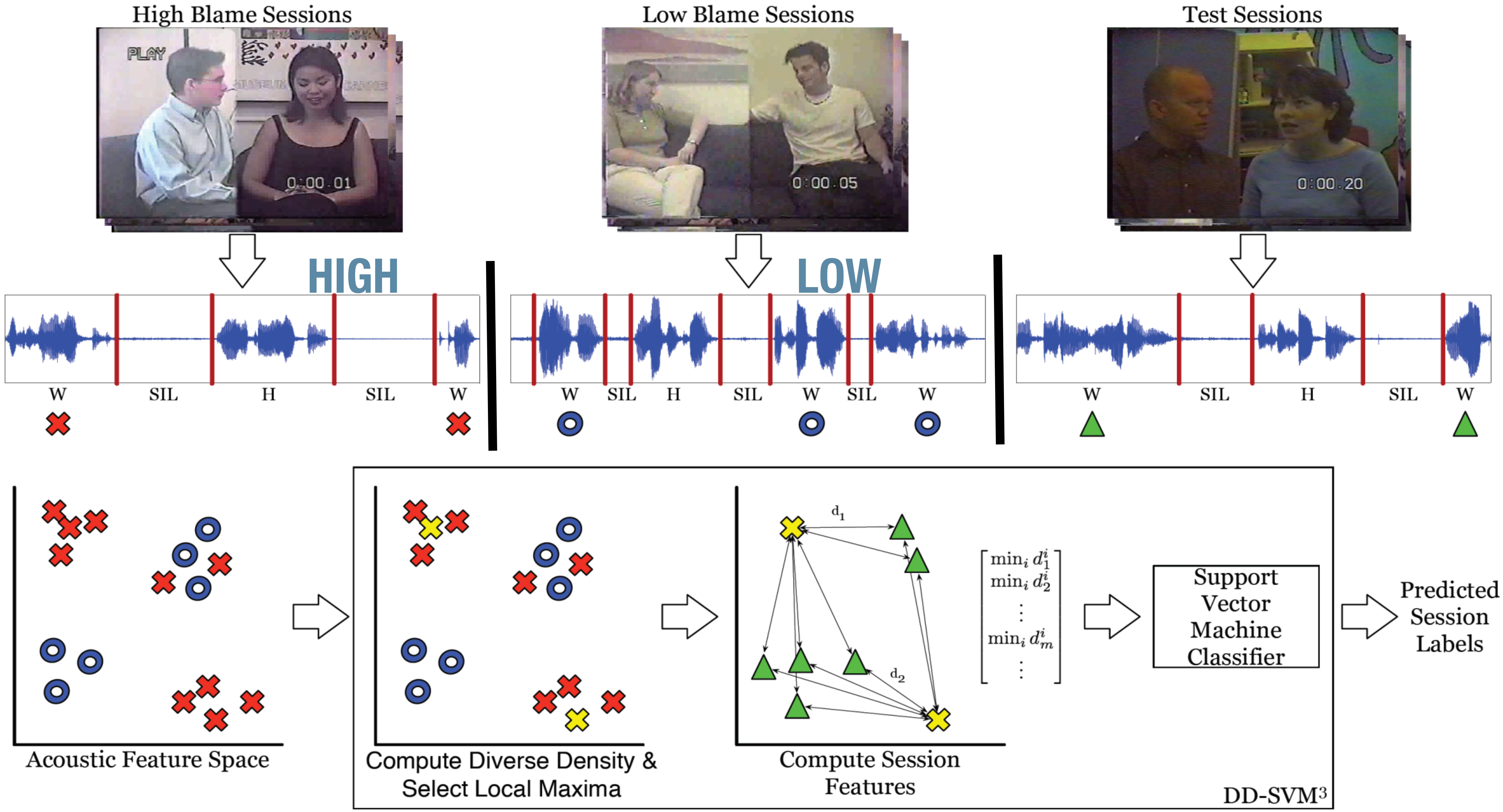
KATSAMANIS, GIBSON, BLACK, NARAYANAN, MULTIPLE INSTANCE LEARNING FOR CLASSIFICATION OF HUMAN BEHAVIOR OBSERVATIONS, ACII 2011
JAMES GIBSON, ATHANASIOS KATSAMANIS, FRANCISCO ROMERO, BO XIAO, PANAYIOTIS GEORGIOU, SHRIKANTH NARAYANAN. MULTIPLE INSTANCE LEARNING FOR BEHAVIORAL CODING. IEEE TRANSACTIONS ON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

Saliency estimation overview



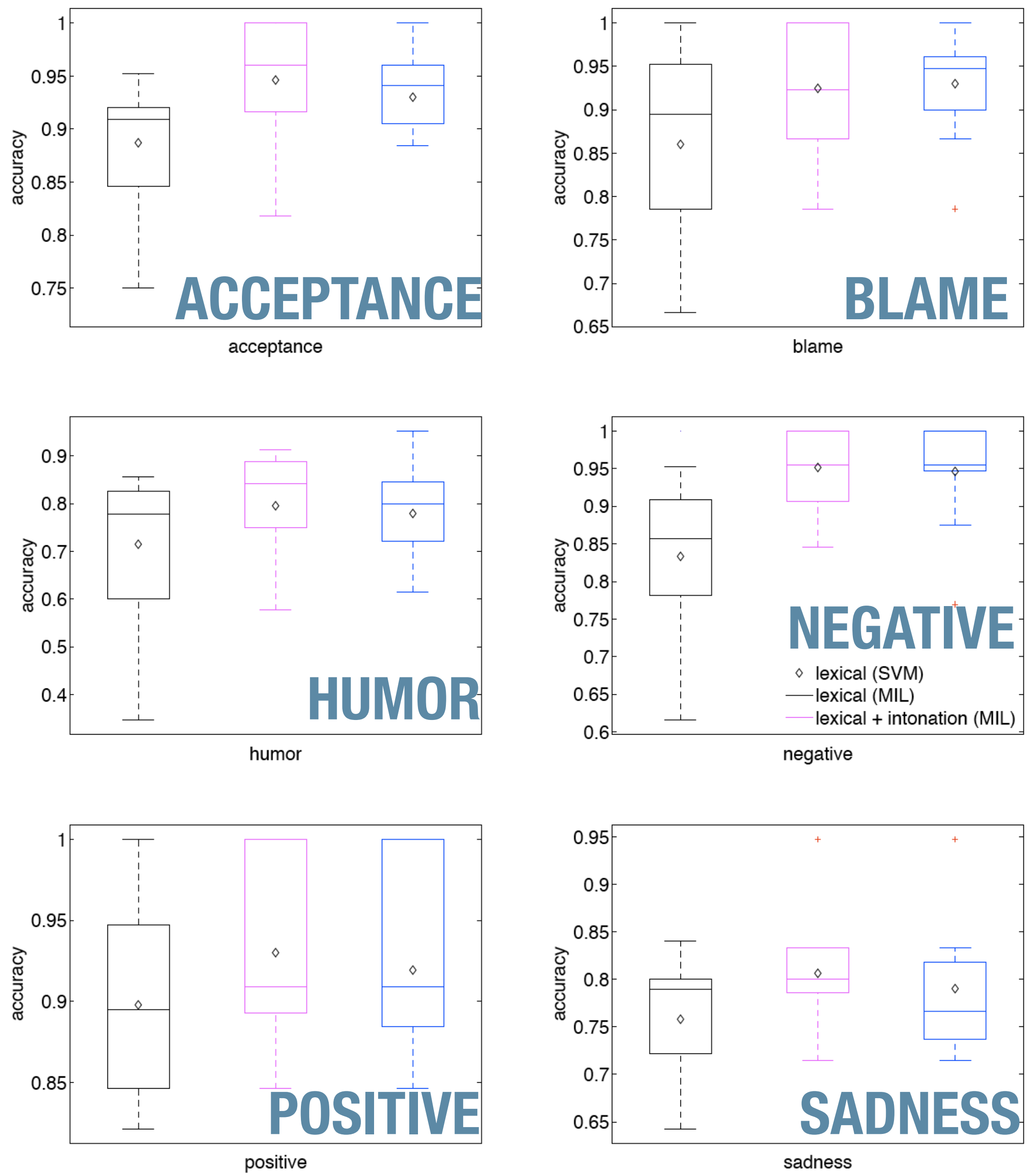
Saliency Detection with Multiple Instance Learning

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

Classification Results



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

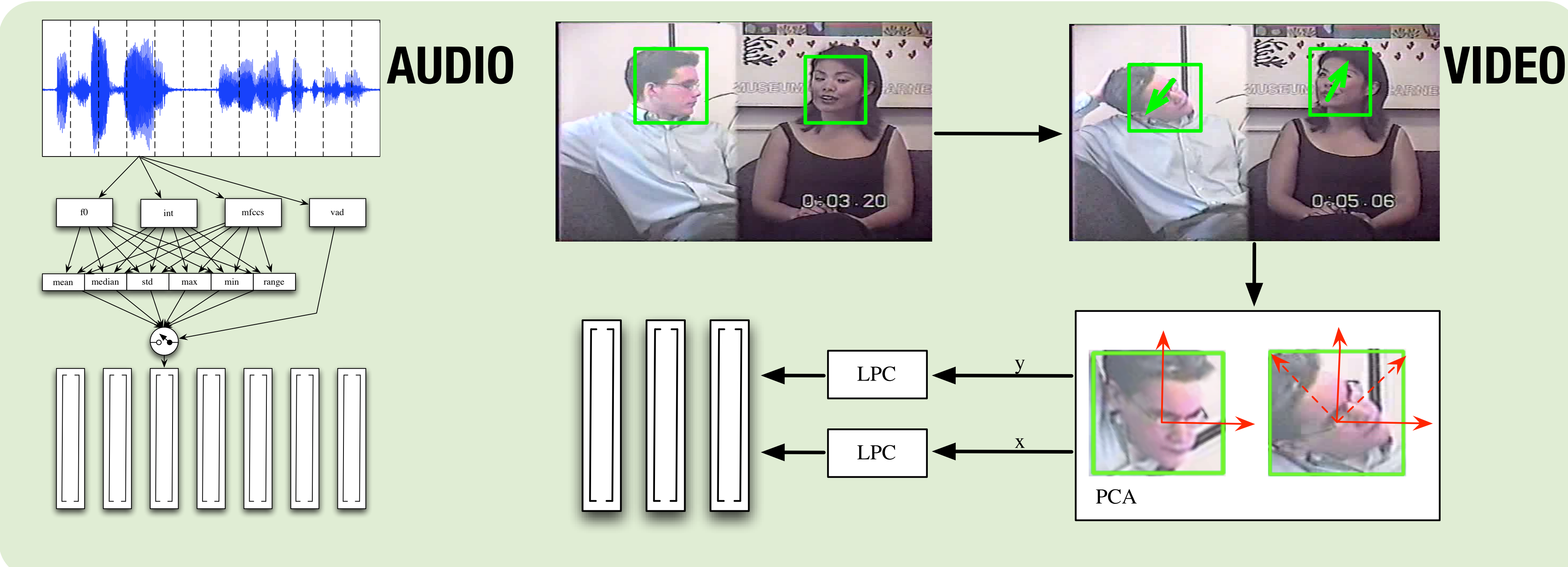
black boxes — baseline: Bag-of-words representation of the whole session (without exploiting saliency estimates)

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

Some technical challenges & 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**

Interaction Models

Interaction Synchrony / Entrainment [Kimura 2006]

Mutual adaptation of verbal/nonverbal behaviors in dyadic interactions

Positive vs. Negative valence in interactions

Higher degree of entrainment in positive interactions [Kimura 2006, Warner 1987]

Entrainment measures as features for automatic classification [Margolin 1998]

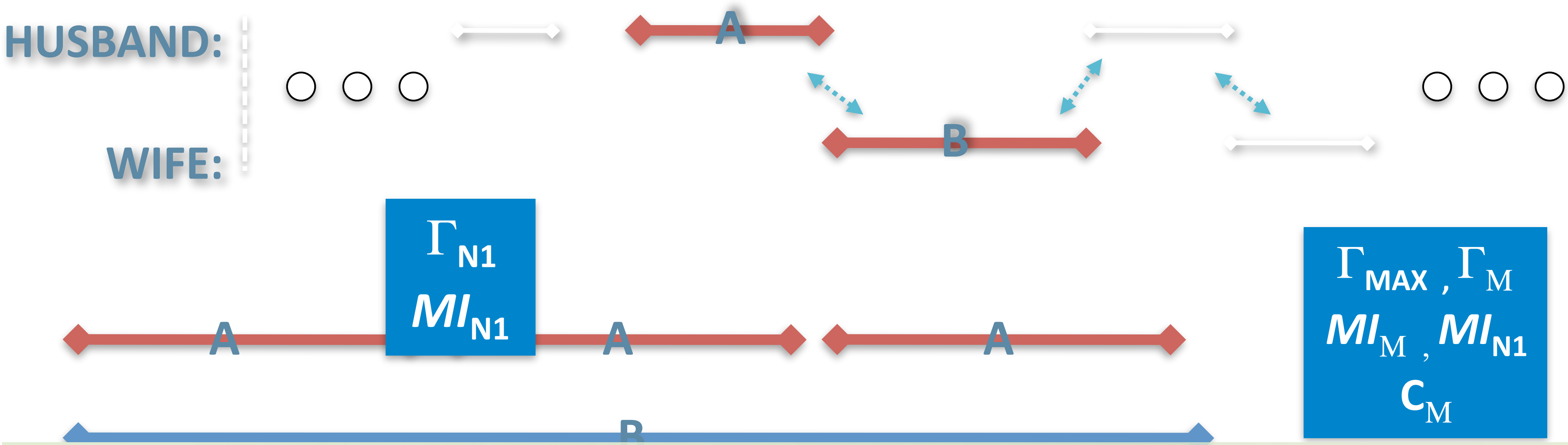
Quantification of Prosodic Entrainment

Signal-derived quantitative measure

"HOW DO TWO PEOPLE SOUND ALIKE AS THEY INTERACT IN A CONVERSATION?"

Plausible Entrainment Measures

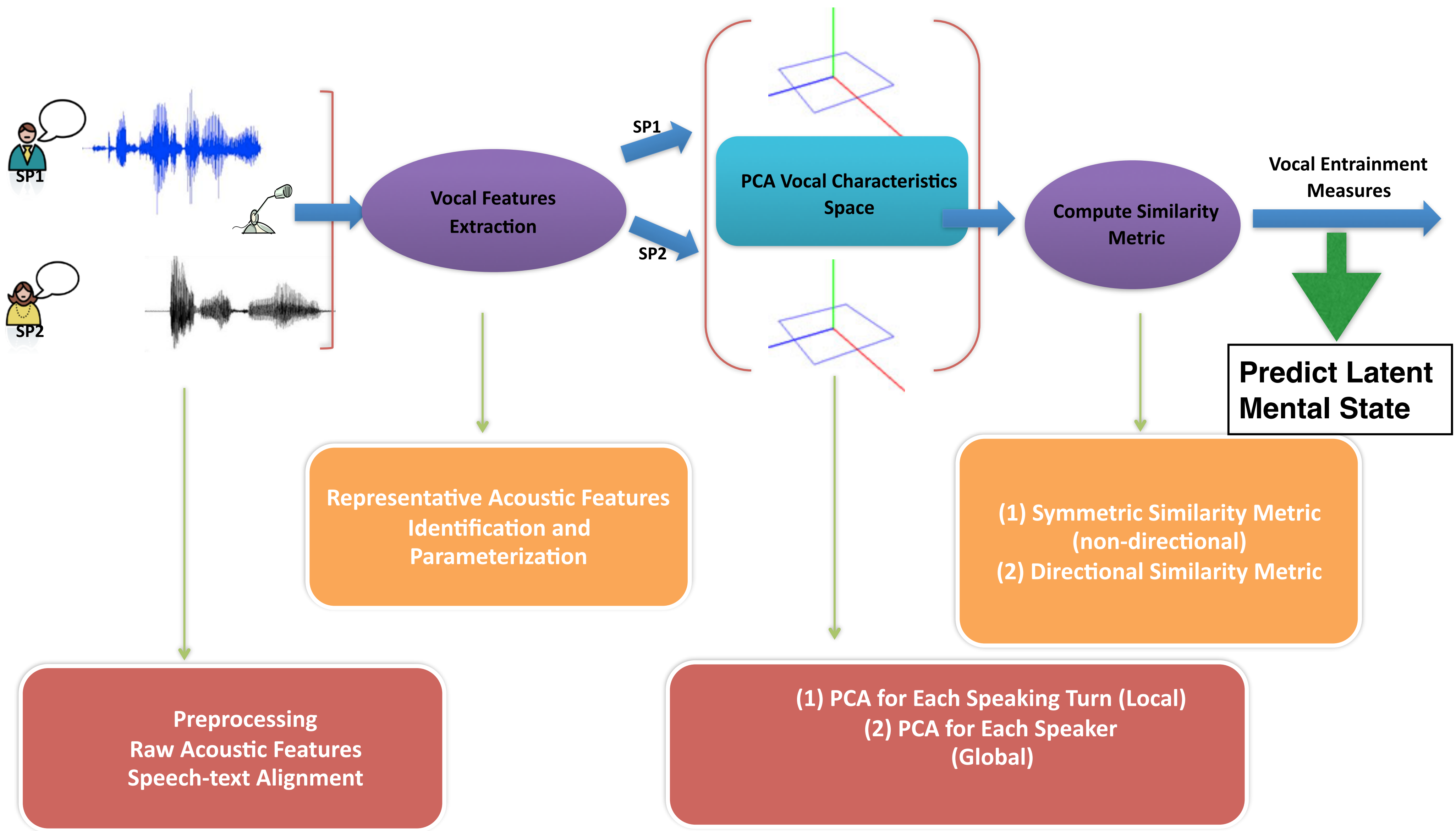
- **Square of correlation coefficients (Γ_{μ}, Γ_{N1}, Γ_{MAX})**
  - Linear dependency between two random variables
- **Mutual information (*mi_{μ}*, *mi_{n1}*, *mi_{max}*)**
  - Information theoretic mutual dependency measure between two random variables
- **Coherence (c_{μ}): Mean spectral coherence across all frequency bands**
  - Commonly used as a measure of degree of causality between a system's input and output relationship



CHI-CHUN LEE, MATTHEW BLACK, ATHANASIOS KATSAMANIS, ADAM LAMMERT, BRIAN BAUCOM, ANDREW CHRISTENSEN, PANAYIOTIS G. GEORGIOU, SHRIKANTH NARAYANAN. QUANTIFICATION OF PROSODIC ENTRAINMENT IN AFFECTIVE SPONTANEOUS SPOKEN INTERACTIONS OF MARRIED COUPLES. IN PROCEEDINGS OF INTERSPEECH, 2011

Computing Vocal Entrainment: A novel measure

"HOW MUCH DO TWO PEOPLE *SYNCHRONIZE* IN A CONVERSATION?"



CHI-CHUN LEE, ATHANASIOS KATSAMANIS, MATTHEW P BLACK, BRIAN R BAUCOM, ANDREW CHRISTENSEN, PANAYIOTIS G GEORGIOU AND SHRIKANTH S NARAYANAN. 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

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, positive 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, ANGELIKI METALLINOU, ENGIN ERZIN, SHRIKANTH NARAYANAN. ANALYSIS OF INTERACTION ATTITUDES USING DATA-DRIVEN HAND GESTURE PHRASES. IN PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON AUDIO, SPEECH AND SIGNAL PROCESSING, 2014

BO XIAO, PANAYIOTIS G. GEORGIOU, ZAC E. IMEL, DAVID C. ATKINS, SHRIKANTH S. NARAYANAN. MODELING THERAPIST EMPATHY AND VOCAL ENTRAINMENT IN DRUG ADDICTION COUNSELING. IN PROCEEDINGS OF INTERSPEECH, 2013

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

TALK OUTLINE

Some behavioral informatics building blocks

- Focus on multimodal data processing
- Affective Computing as an example

Some Case Studies

- Dyadic interaction of distressed couples
  - Marital therapy
- ✓ Autism Spectrum Disorders
  - Quantifying social interaction and communication: behavior stratification: supporting diagnostics, evaluating outcomes
- Addiction
  - Understanding and evaluating psychotherapy

Autism Spectrum Disorders (ASD)

- **1 in 68 US children diagnosed with ASD (CDC, 2014)**
- **ASD characterized by**
  - Difficulties in social communication, reciprocity
  - Repetitive or stereotyped behaviors and interests
- **Technology possibilities in ASD include**

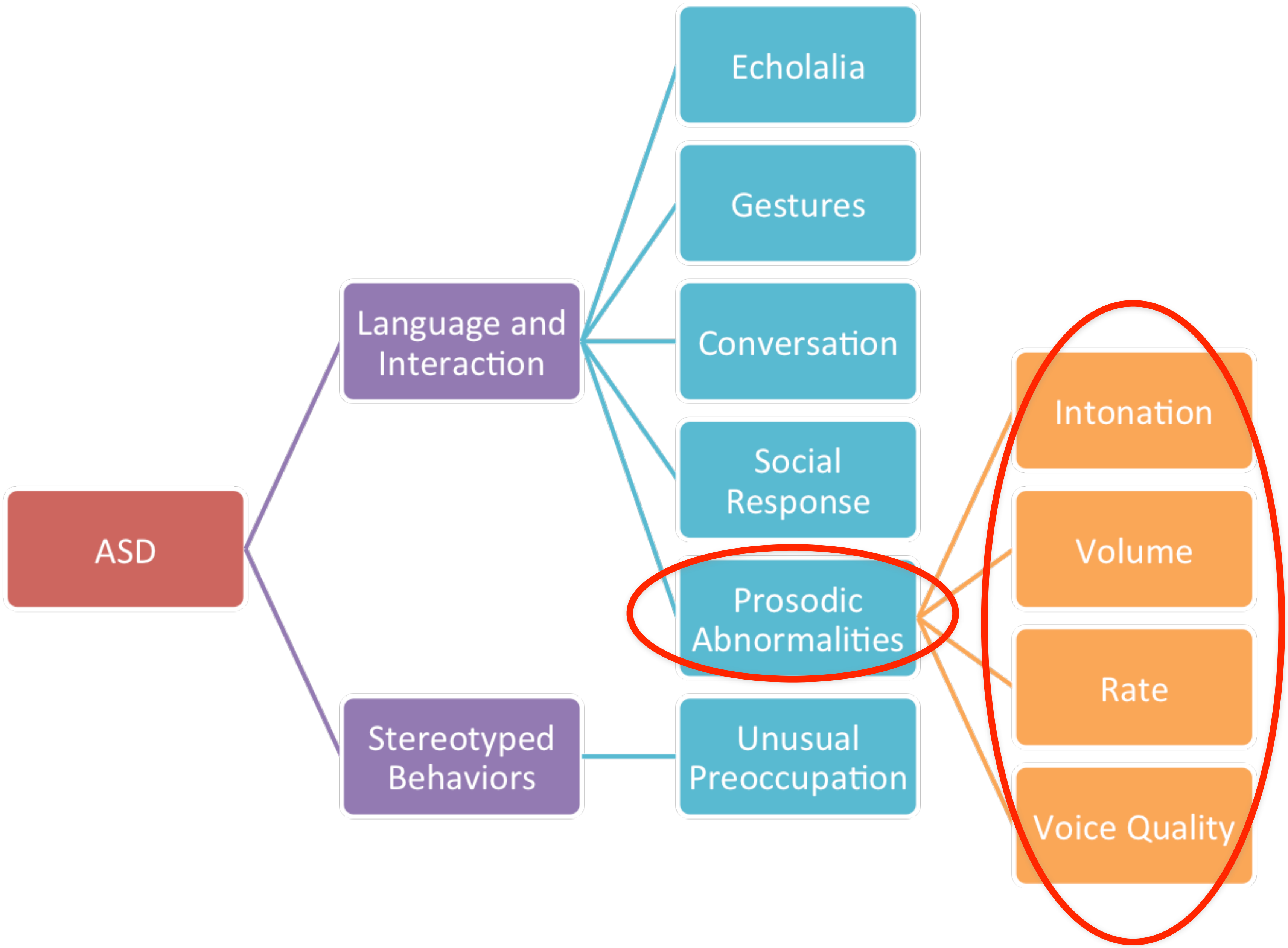
  Computational techniques to
  - Better understand communication and social patterns of children
  - Stratify phenotyping with quantifiable and adaptable metrics
  - Track, quantify children's progress during interventions

  Interfaces/systems to elicit, encourage, analyze behavior
  - Complex, but phased; Structured; Naturalistic

Analyzing Interaction in ASD

- Assessment, Intervention, Game play/training Examples

ASD Assessment



Quantifying Atypical Prosody

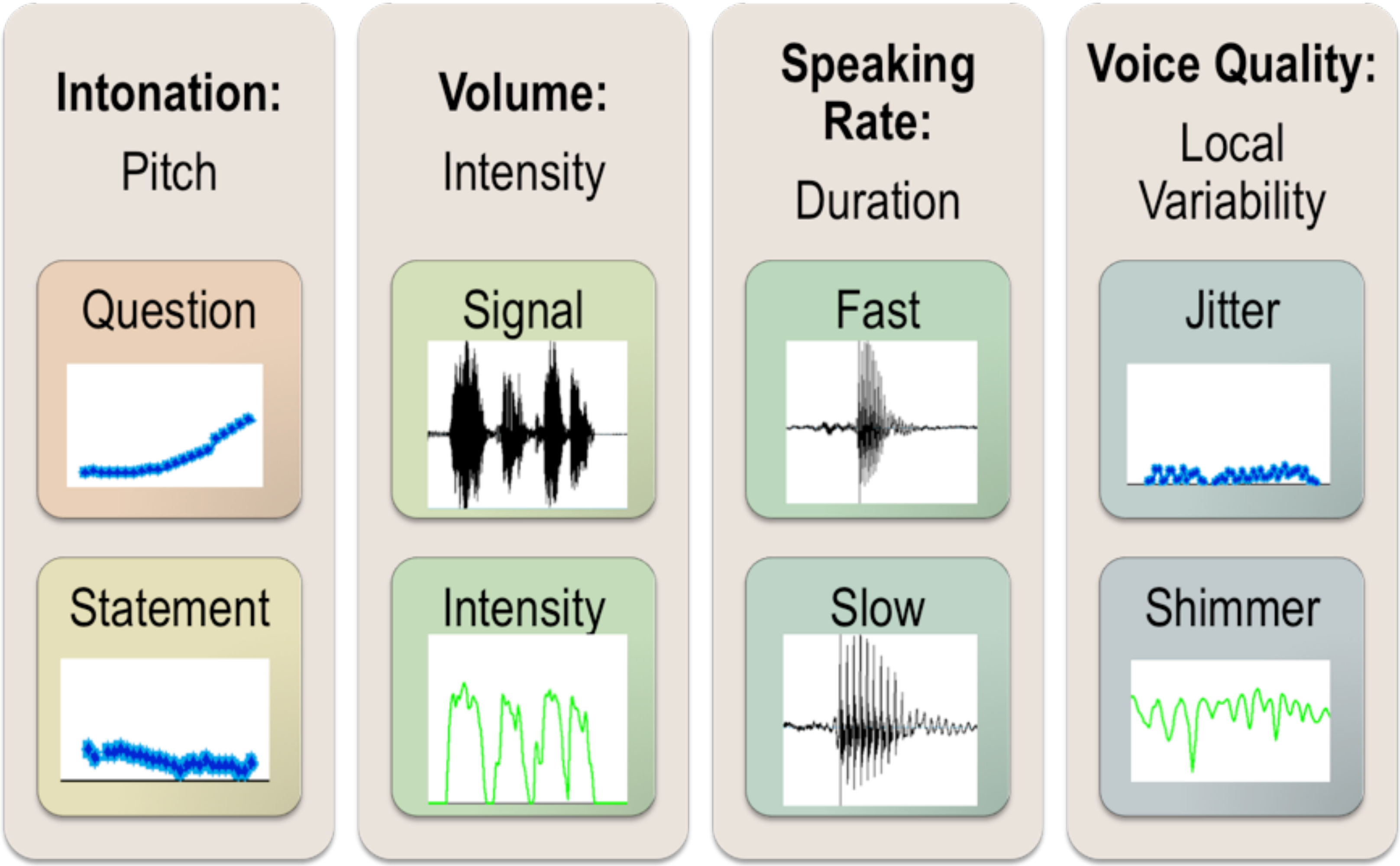
Qualitative descriptions are general and contrasting

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



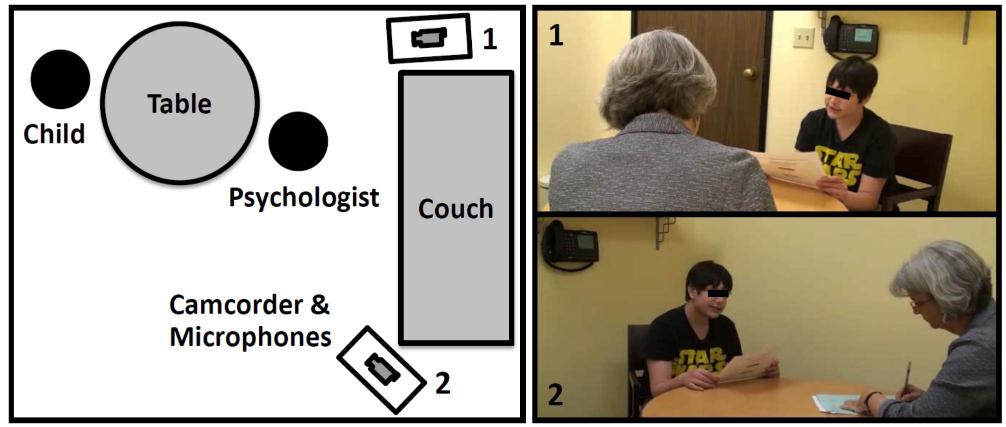
- 24 Features: pitch (6), volume (6), rate (4), and voice quality (8)
  - Intonation: F0 curvature, slope, center
  - Volume: Intensity curvature, slope, center
  - Rate: Boundary (turn end word), Non boundary
  - Voice Quality: Jitter, Shimmer, CPP, HNR

✦ ***median, IQR of above***

Our Case study Setup

Approach

- **Automatic measures from spontaneous speech**
  - Create generally applicable tools for discovery
- **Data**
- **N=28 children.**
- **ADOS module 3 Interviews**
  - USC CARE Corpus



Hypotheses

1. **Children with ASD will demonstrate correlation between acoustic-prosodic cues and severity of ASD-related impairment**
2. **Psychologist's speech is also informative of rated severity (both participant and evaluator)**

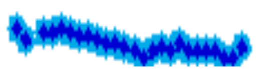
MATTHEW P. BLACK, DANIEL BONE, MARIAN E. WILLIAMS, PHILLIP GORRINDO, PAT LEVITT, AND SHRIKANTH NARAYANAN. THE USC CARE CORPUS: CHILD-PSYCHOLOGIST INTERACTIONS OF CHILDREN WITH AUTISM SPECTRUM DISORDERS. PROCEEDINGS OF INTERSPEECH, 2011.

Atypical Prosody & Interaction

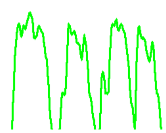
Spearman's Correlation between rated severity and prosodic cues

Child's Prosody

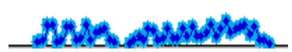
- "Monotone"

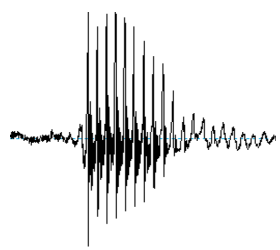
  p<0.01
- "Abnormal volume"

  p<0.05
- "Breathy/Rough"

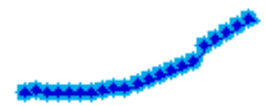
  p<0.01
- Slower speaking rate

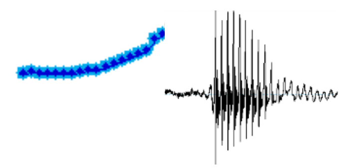
  p<0.05

Psychologist's Prosody

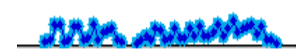
- Questions/affect

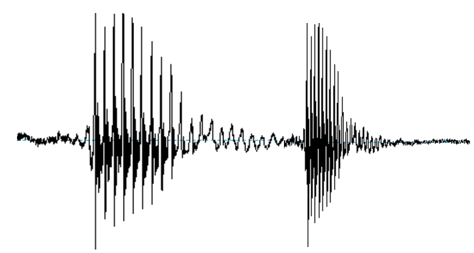
  p<0.05
- Variable prosody

  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.

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

Quantifying Qualitative Social Perceptions

Atypicality in Facial Expressions of ASD children

Understanding the expression and perception of social cues in ASD: What makes the difference?

Example: Production of Affective Facial Expressions During Smile Imitation Task

Computational Targets

Quantify atypicality of smile

Region-based activation

Synchrony & symmetry

Tanaya Guha, Zhaojun Yang, Anil Ramakrishna, Ruth Grossman, Darren Hedley, Sungbok Lee, Shrikanth Narayanan. ON QUANTIFYING FACIAL EXPRESSION-RELATED ATYPICALITY OF CHILDREN WITH AUTISM SPECTRUM DISORDER. In ICASSP, 2015

Angeliki Metallinou, Ruth Grossman, Shrikanth Narayanan. Quantifying Atypicality In Affective Facial Expressions Of Children With Autism Spectrum Disorders. In Proceedings of the IEEE International Conference on Multimedia & Expo (ICME), 2013

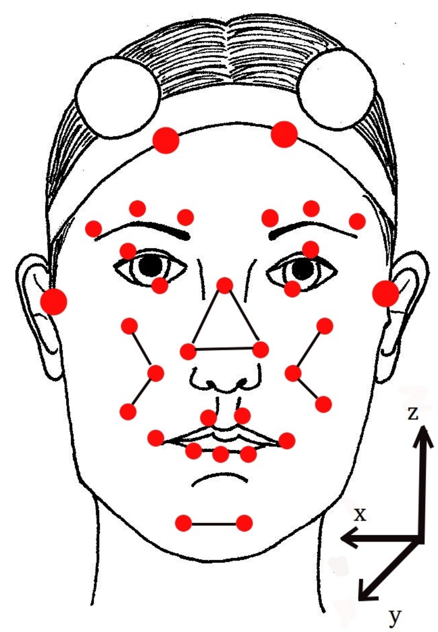
ASD and facial expressions

- ASD linked to production of atypical facial expressions and prosody [Asperger, 1944] [Kanner 1968]
- Asynchrony in coordinating speech and gestures [DeMarchena etal., 2010]
- Facial expressions often perceived as ‘atypical’ or awkward [Grossman etal., 2012][Yirmiya etal., 1989]
    - *Awkwardness impression is hard to quantify*

- Use MoCap technology and statistical methods to **computationally quantify**:
    - What causes this impression of awkwardness?
    - What differentiates neurotypical from ASD populations?
- Gain insights that can’t be obtained otherwise

Methods and Findings

- Motion Capture (MoCap)
  - Detailed capture of facial expressions
- Functional Data Analysis (FDA) and statistical testing
  - Model, analyze, discover properties
- Multidimensional Scaling
  - Visualization of subject variability and similarity



- Findings
  - Increased **facial motion asynchrony** for ASD
  - Increased **facial motion roughness**
  - Consistently **greater expression variability**
    - Idiosyncratic face gestures

Angeliki Metallinou, Ruth Grossman, Shrikanth Narayanan. Quantifying Atypicality In Affective Facial Expressions Of Children With Autism Spectrum Disorders. In Proceedings of the IEEE International Conference on Multimedia & Expo (ICME), San Jose, CA, 2013

Interventions for Addiction

- Motivational Interviewing: Assessment, Training
- Characterizing empathic behaviors

Psychotherapy: Addiction

Motivational Interviewing: Widely used

- Client's (interviewee) own will of making a change
- Therapist (interviewer): understand, facilitate, do not dictate
- Goal-oriented, highly-structured
- Non-confrontational, non-judgmental, dialog setting

COMPUTATIONAL BEHAVIOR MODELING POSSIBILITIES

Interview efficacy: Modeling constructs such as “reflections”

Client-counselor Interaction dynamics: Empathic behavior

- Computational modeling: insights into the expressed empathy
- Use speech, spoken language, nonverbal cues
- Data from several clinical intervention studies, coded by experts

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 (ALSO INTERSPEECH 2012).

Addiction Psychotherapy Corpora

- Originally collected for psychotherapy process research
- ~800 audio/video sessions from 5 different, brief intervention studies:
    - HMCBI, ESP21, ESPSB, iCHAMP, ARC
- ~10% (155 sessions) manually transcribed and annotated by trained coders
    - Utterance Level Behavioral Codes (MISC: Motivational Interviewing Skills Code)
    - Session Level Behavioral Codes (MISC, MITI Motivational Interviewing Training Integrity)
    - Outcomes

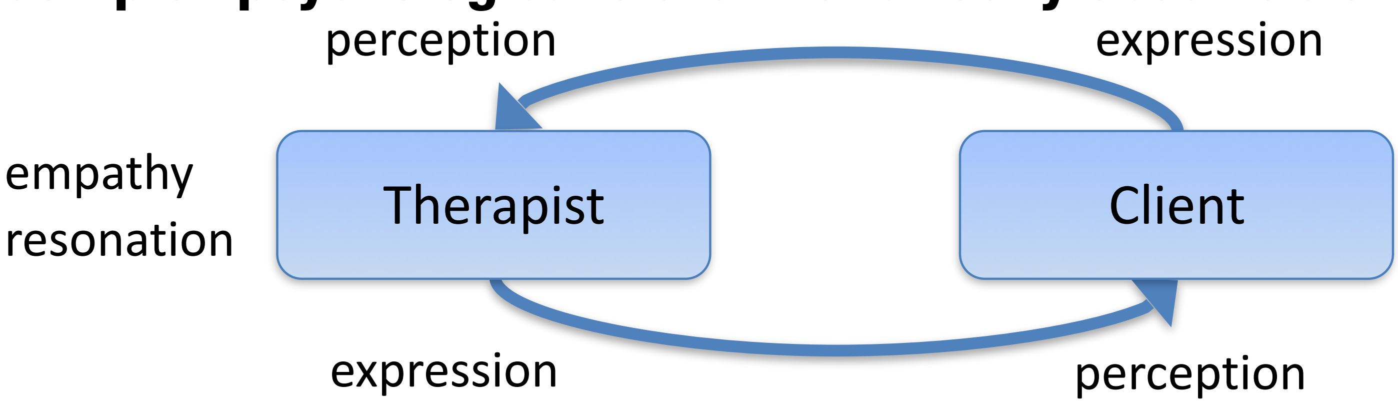
Motivational Interviewing

Sample (training) video

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

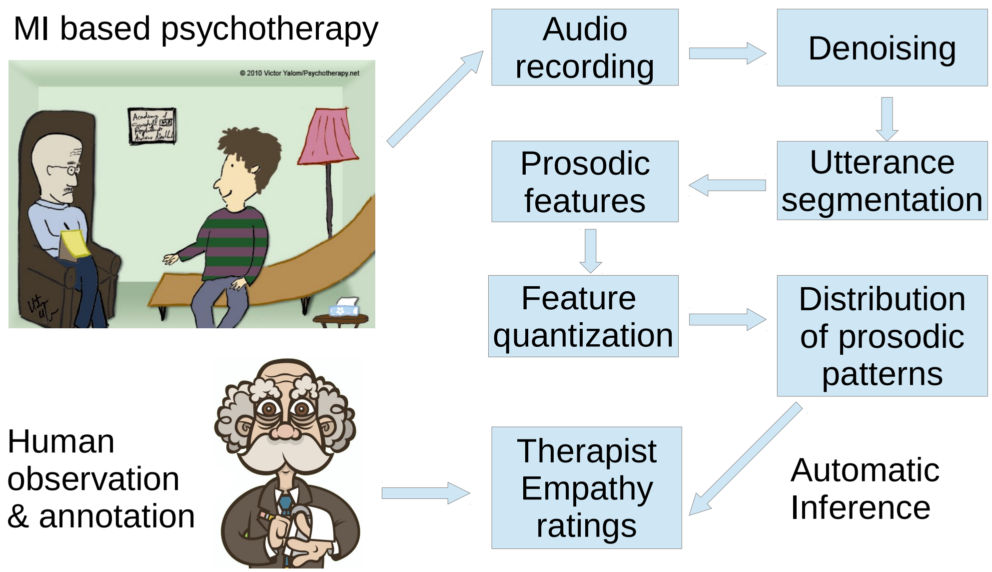
Empathy & psychotherapy

- **Definition**
  - Emotional simulation
  - Perspective taking
  - Emotion regulation
- **Evaluation**
  - Key performance index
  - Explain outcome
- **Addiction counseling**
  - **Summative assessment of behaviors in an interval**
  - **Manual observational evaluation not scalable**
  - **Complex psychological state: not directly observable**



Modeling Expressed Empathy

- **Speech prosody and empathy: neurological and behavioral evidence of links**



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

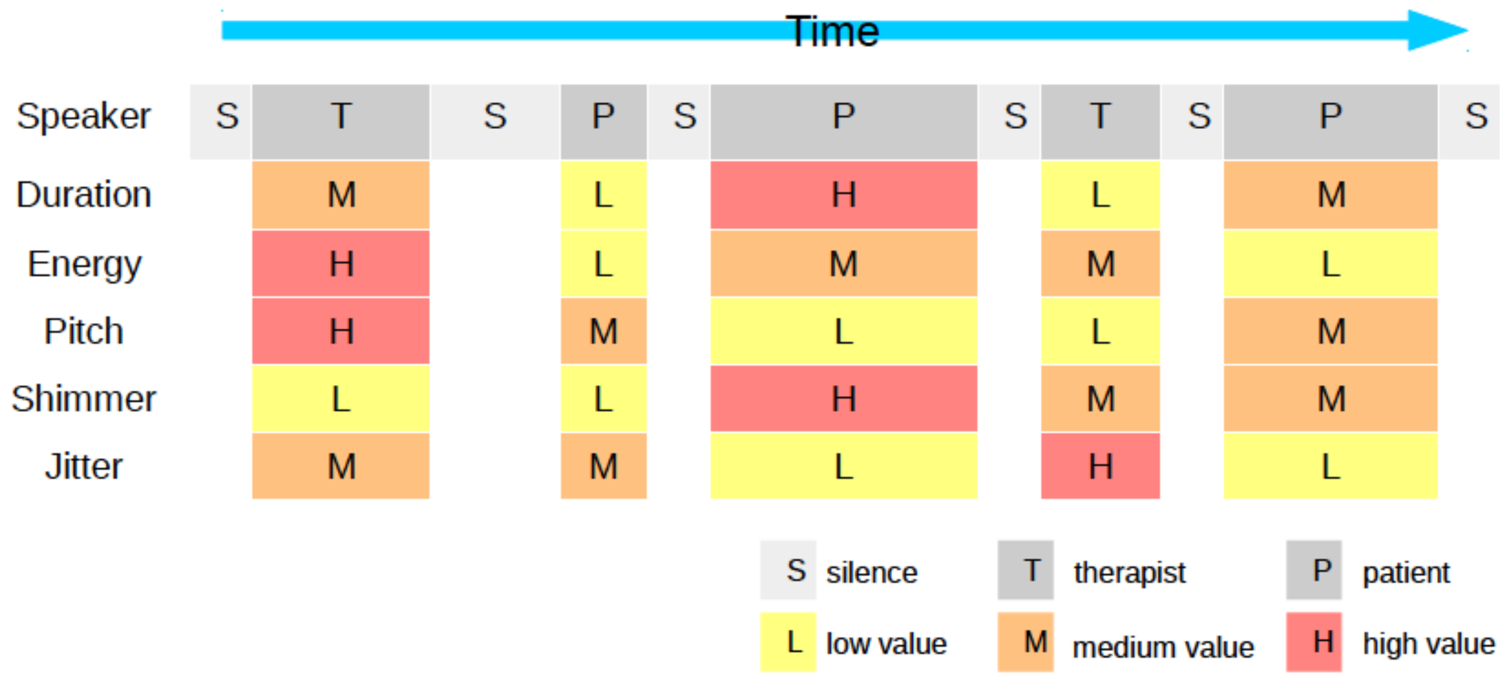
Key Findings

- Prosodic correlates of perceived therapist empathy
- Quantization & joint modeling of prosody derives salient prosodic patterns

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

Computational Modeling of Empathic Behavior

- **Speech prosody measures**
  - Extract, quantify, and model the distribution of prosodic cues
  - *Quantized features*: turn duration, energy, pitch, jitter, shimmer

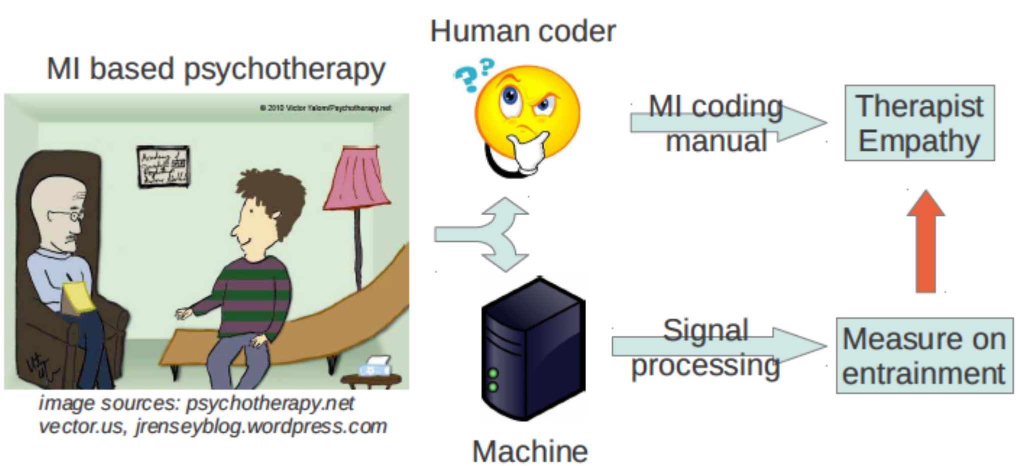


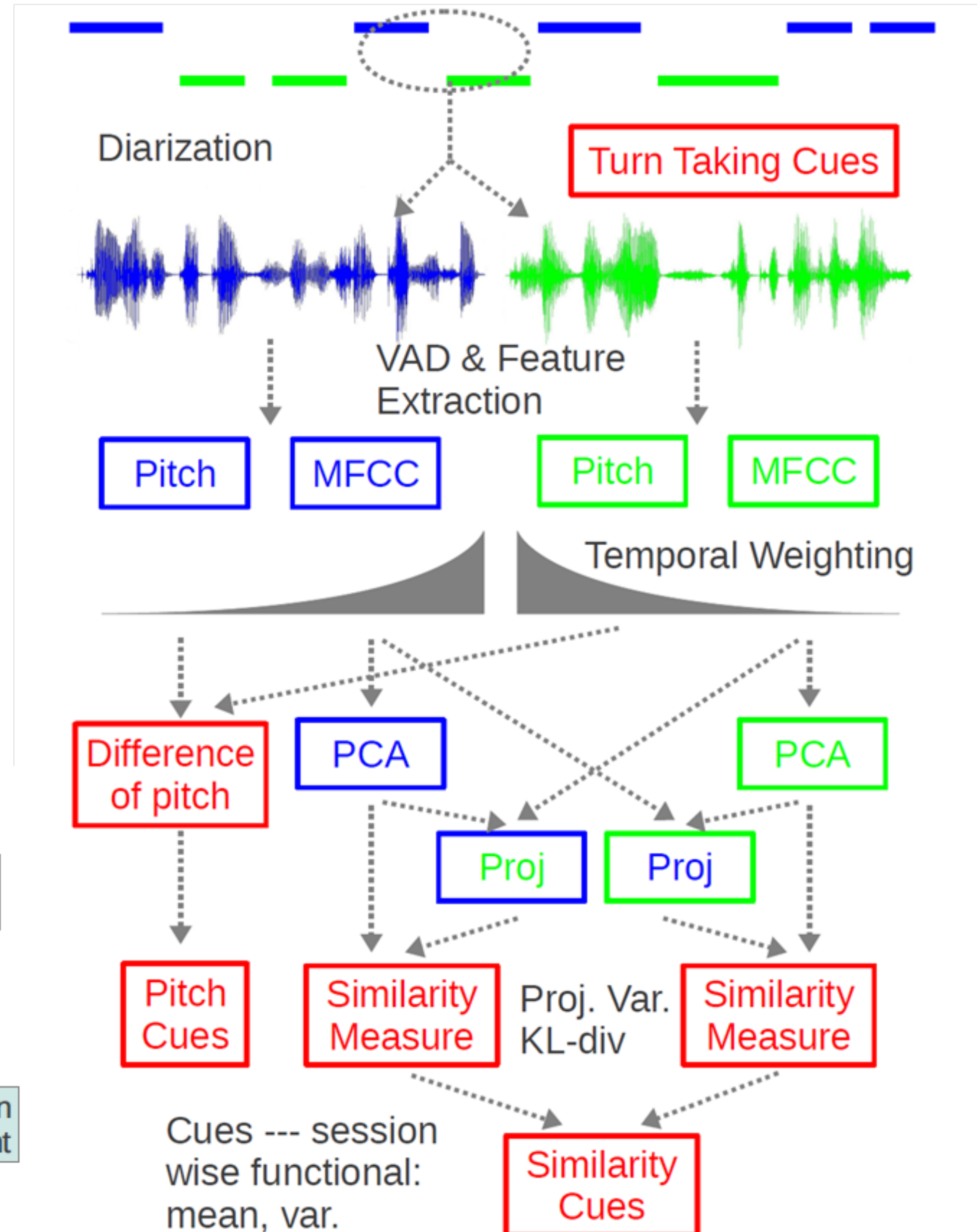
- ***Results*: lower perceived empathy of therapist when:**
  - Therapist has higher energy values
  - Therapist has higher pitch values

Bo Xiao, *et al.*, Modeling Therapist Empathy through Prosody in Drug Addiction Counseling, 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**

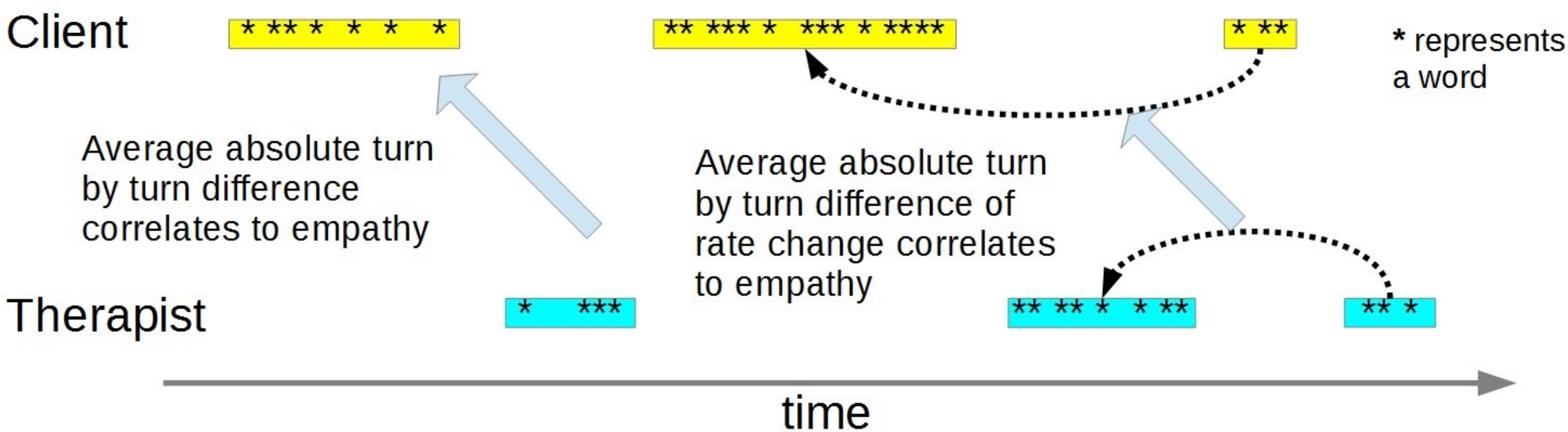




Bo Xiao, et al., Modeling Therapist Empathy and Vocal Entrainment in Drug Addiction Counseling Proceedings of Interspeech, 2013

Speech rate entrainment

- Difference of average speech rates NOT correlated
- Turn-by-turn difference of speech rates, and difference of rate accelerations
  - Correlation in range -0.2 to -0.3, p<1e-3
  - Robust to alignment error



Speech rate entrainment

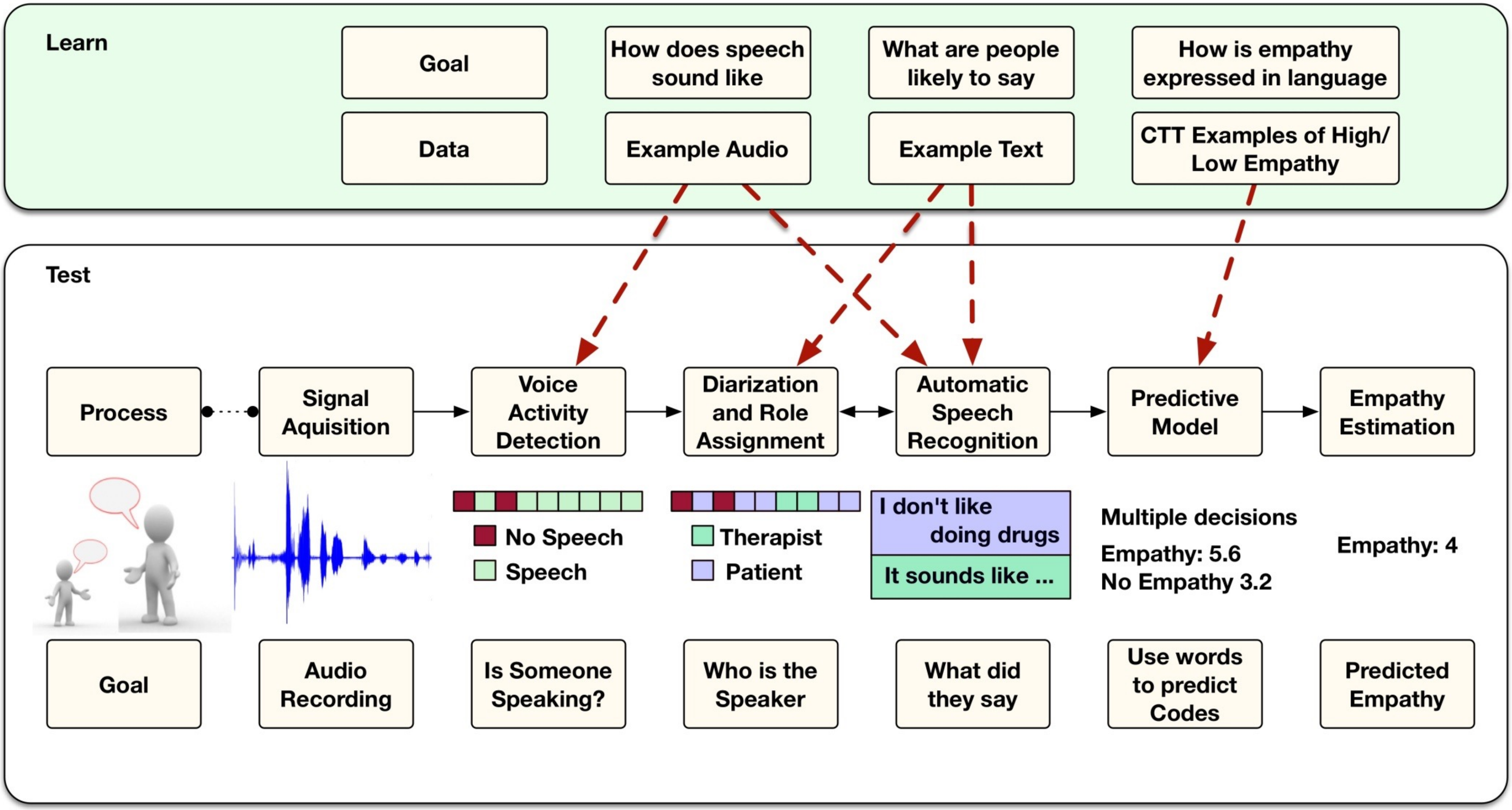
Turn Taking Cues

Consider ratio by time or segment count

Turn Taking Cues	Positive / Negative	Correlation to empathy	p-value
Ratio of client speech and pause	Positive	0.3	1e-3
Ratio of therapist speech and pause	Negative	-0.3	1e-3
Ratio of pauses (intra-speaker) (only segment)	Positive	0.2	0.01
Ratio of gaps (inter-speaker)	Negative	-0.2	0.01

“Sound to code” system:

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

Language of Therapy: Psycholinguistic Norms

- Linguistic norms are numerical ratings which reflect the similarity of a particular word to various categories
- Psycholinguistic norms represent word relations to psychological processes such as affect

norm	description
age of aquisition	expected age at which the word is aquired
arousal	degree of excitement (versus calmness)
context availability	number of contexts in which the word appears
concreteness	degree of concreteness (versus abstractness)
dominance	degree of control over a situation
familiarity	how commonly a word is expressed
gender ladenness	degree of feminity (versus masculinity)
imageability	degree of ease in forming a mental image
meaningfullness	how associated a word is to other words
pleasantness	degree to which pleasant feeling are associated
pronounceability	degree of ease in pronouncing the word
valence	degree of emotional positivity (versus negativity)

Psycholinguistic Norm Features

- Each word receives a raw score according to 13 psycholinguistic dimensions and 47 part of speech tags (POS)
  - POS tags are from the Penn Treebank [Marcus, et al. 1993]
- 9 functionals features are computed across all the words of a given speaker in each session with 2 normalization schemes
  - length (number of tokens), min, max, extremum (value furthest from zero), sum, average, range, standard deviation, variance
  - normalized and unnormalized
- *Psycholinguistic Norm Features* (PNF)
  - comprised of 10,998 feature dimensions[Malandrakis, et al. 2015]

Nikolaos Malandrakis and Shrikanth Narayanan. Therapy Language Analysis using Automatically Generated Psycholinguistic Norms. In Proceedings of Interspeech, 2015

Predicting Empathy - Fusion

- **fusion: mean posterior score**

LIWC: Linguistic Inquiry and Word Count
PNF: psycholinguistic norm features

features	UAR	Spearman's ρ
ngram+LIWC	70.72	0.5606
ngram+PNF	71.44	0.6023
LIWC+PNF	69.20	0.5202
all-late	75.28	0.5952

James Gibson, Nikolaos Malandrakis, Francisco Romero, David Atkins and Shrikanth Narayanan. Predicting Therapist Empathy in Motivational Interviews using Language Features Inspired by Psycholinguistic Norms. In Proceedings of Interspeech, 2015

Insights: Feature Analysis

- “it sounds like…”
- hearing and perception - reflections
- abstract, unambiguous language

feature set	top features
unigram	sounds (0.32), ever (-0.31), severe (0.03), meds (0.28)
bigram	sounds like (0.33), to ten (0.33), severe risk (0.32), drug abuse (0.28)
trigram	it sounds like (0.31), in about a (0.30), abuse screening test (0.32), zero to ten (0.26)
LIWC	hear (0.36), perceptual (0.35), anxiety (0.30), affect (0.28)
PNF - content words	meaningfulness (-0.37), concreteness (-0.31), imageability (-0.28), context availability (-0.23)
PNF - verbs	meaningfulness (-0.43), context availability (-0.42), age of aquisition (0.40), pleasantness (-0.35)
PNF - nouns	concreteness (-0.35), imageability (-0.34), meaningfulness (-0.23)

Behavioral signal processing: 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

TALK SUMMARY:
Open Challenges → RICH R&D Opportunities

- Robust capture and processing of multimodal signals
- Capturing natural behavior in ecologically valid ways
- Behavior representations for computing
- Reflecting multiple (diverse) perspectives and subjectivity
- Feature-behavior correspondence: human like processing
- Scientifically and computationally principled modeling
- Reliably characterizing atypical and disordered patterns
- Data provenance, integrity, sharing, and management
- ***Developing productive partnerships between various domain experts and stakeholders***

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 constructs
- **Computational advances: sensing, processing and modeling**
  - Signals and systems approach to human interaction studies
  - Support **BOTH** human and machine decision making
- **Exciting technological and societal possibilities**
  - Opportunities for interdisciplinary and collaborative scholarship
  - Enable broader access, and directly impact directly various walks of life

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
✓ CREATE TOOLS FOR DISCOVERY

SAIL

Work reported represents collaborative efforts with numerous colleagues and collaborators

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

SHRIKANTH NARAYANAN AND PANAYIOTIS GEORGIOU. BEHAVIORAL SIGNAL PROCESSING: DERIVING HUMAN BEHAVIORAL INFORMATICS FROM SPEECH AND LANGUAGE. PROCEEDINGS OF IEEE. 101(5): 1203 - 1233, 2013.