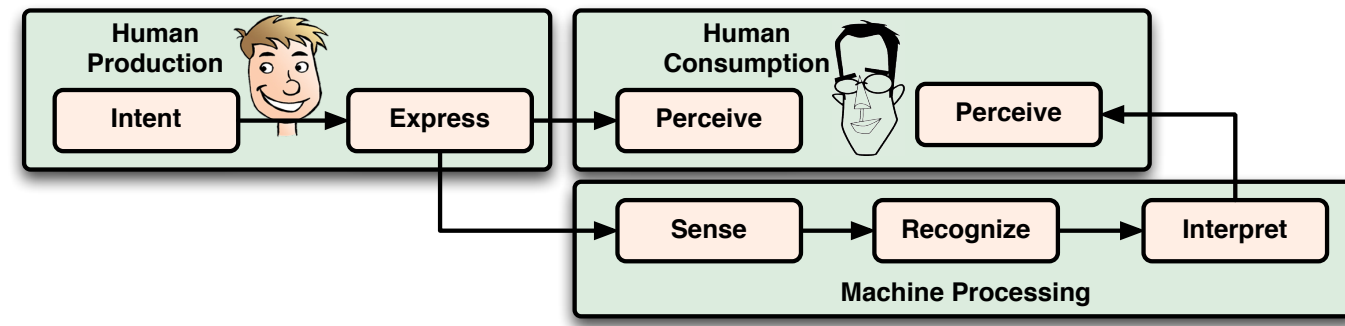


### Abstract

#### Goal:

- Transform observational behavior analysis
- Through computational framework
- Modeling of emotionally-rich human interactions
- Signal processing and machine learning
- Existing family therapy data
- Alleviate the tedium of manual annotation
- Offer new analysis capabilities and empower the mental health experts

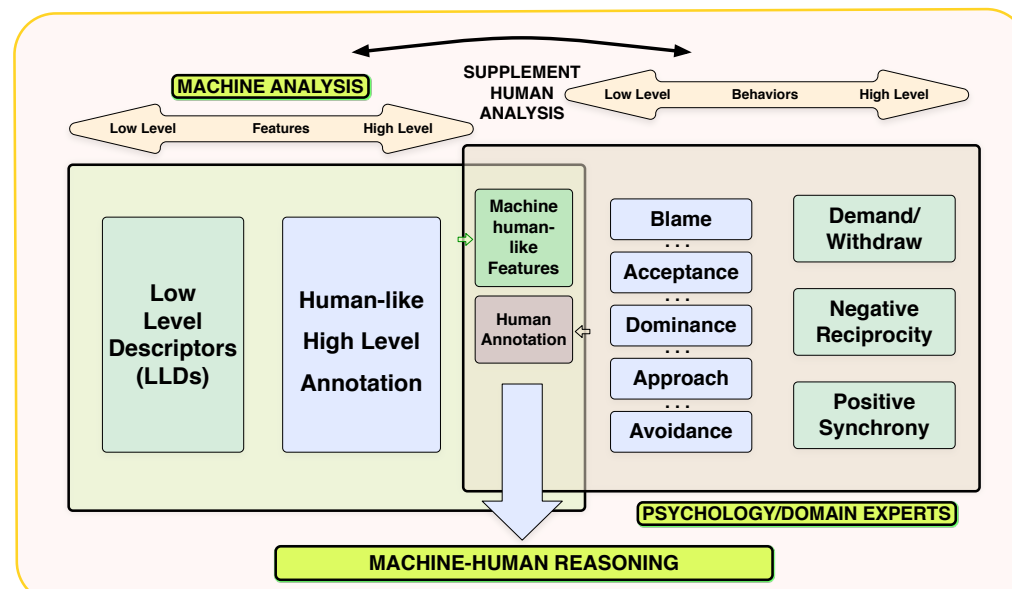


**Significance:** USA-10mil people receive psychotherapy every year and state of the art hasn't changed for decades

### Approaches

+ **This poster:** [- Other two posters]

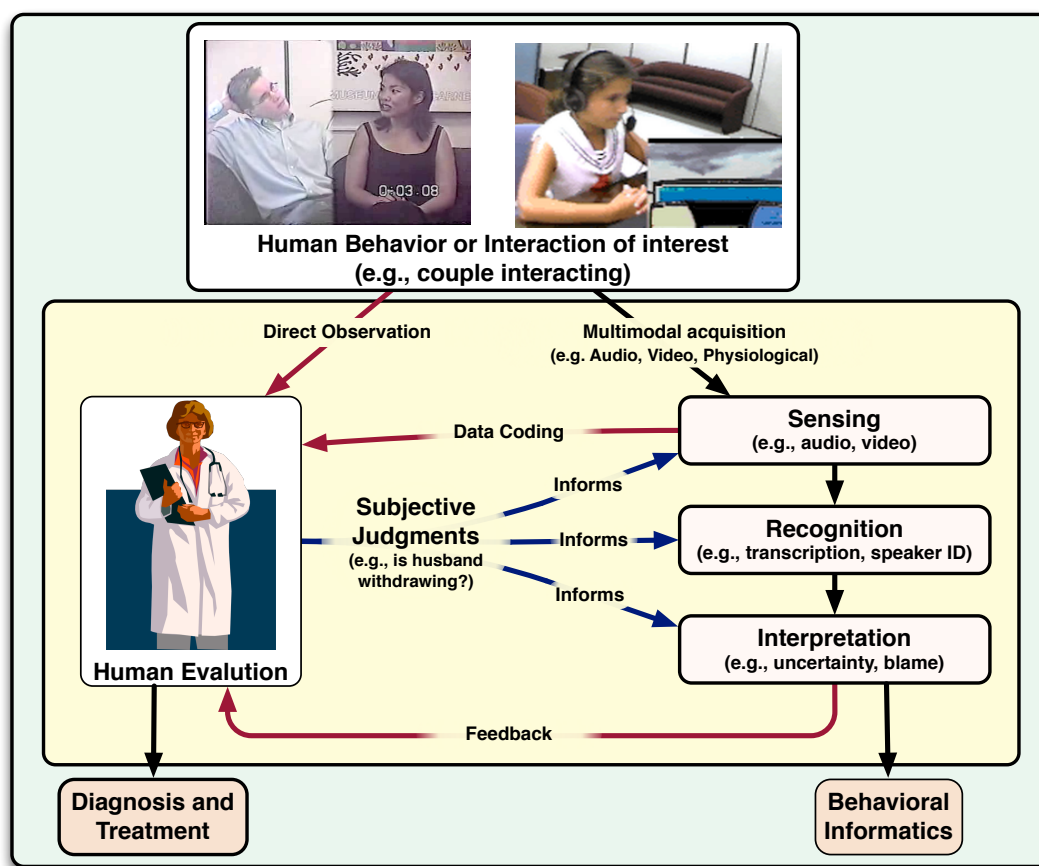
- + Model interlocutors independently:
  - × Lexical, acoustic and visual modalities
- Model dynamics of interlocutors
- Incorporate Saliency



### Data

#### Couple Therapy Corpus

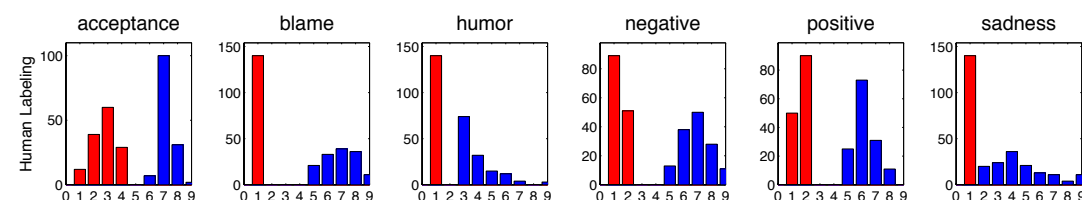
- 117 real distressed couples
- 10-minute dyadic interactions
- 596 sessions (96 hours)



### Data used

#### Audio/Lexical and Visual subsets used

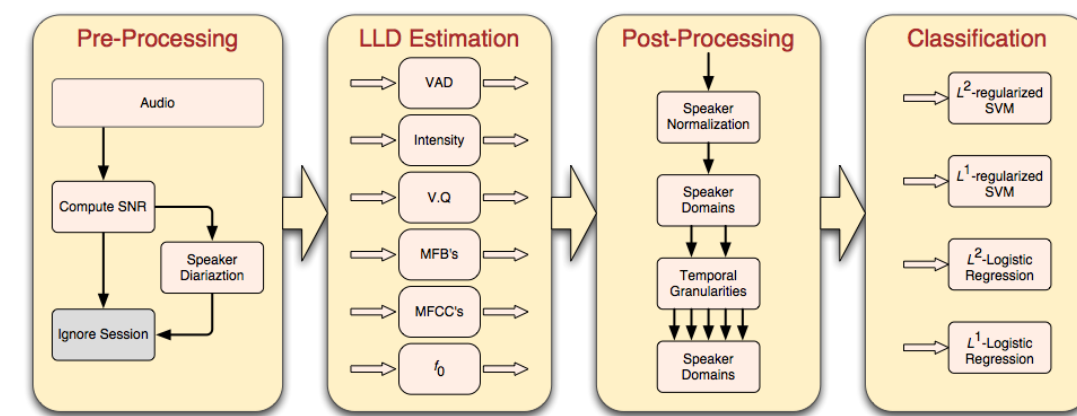
- Use top/bottom 20% for audio, lexical and 25% for video
- Choose subsets with acceptable audio/video qualities
- Used 6 codes with highest human agreement
- Some distributions skewed and not very separable



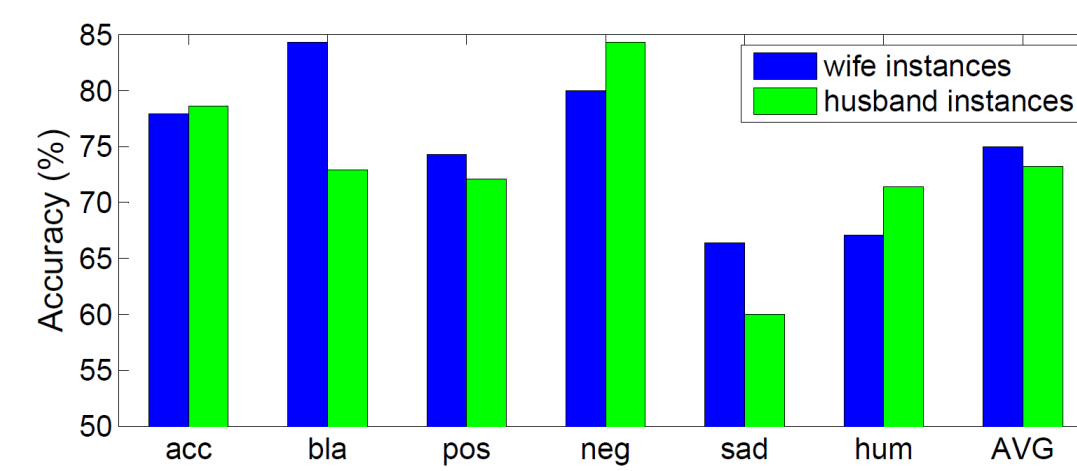
### Acoustic Classification

#### Q: Does acoustic channel capture behavior?

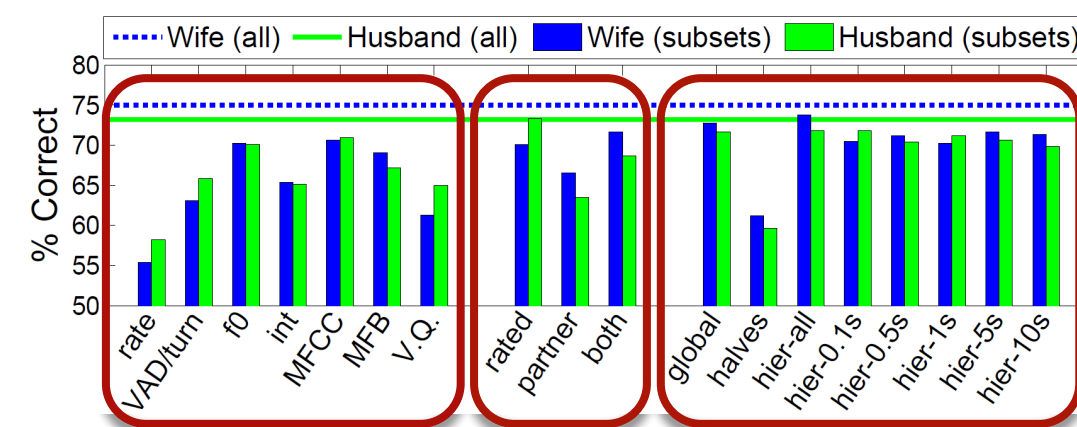
- Frame-level low-level descriptors (LLDs)
  - Prosodic: speech/non-speech, rate, f0, intensity
  - Spectral: 15 MFCCs, 8 MFBs
  - Voice quality: jitter, shimmer
- Separate features for (wife, husband, all)
- 7 temporal granularities
  - Global: entire session
  - Halves: 2nd half - 1st half
  - Hierarchical: 0.1s, 0.5s, 1s, 5s, 10s windows
- 14 static functionals (e.g., mean, std. dev.)



#### Results with logistic regression (L2-regularized)

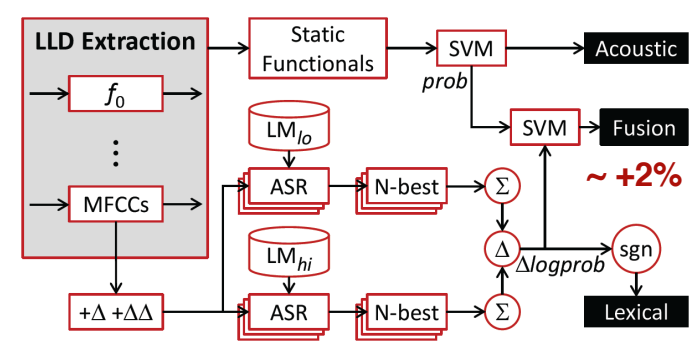


#### Importance of different features



### Fusion

Modalities provide complementary information:



### Lexical

#### Q: Does lexical channel capture behavior?

- Test from reference text
- Test from (unoptimized) ASR output

#### Example Transcript

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

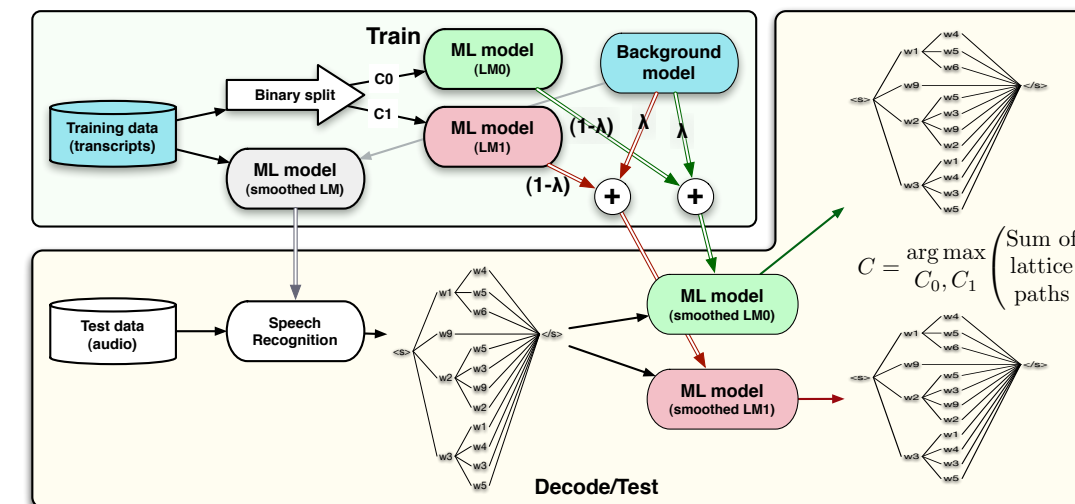
- Train ngram-models (here n=1) in ML manner
- Smooth with UBM
- Score transcripts to classify

#### Oracle classification results

code vs λ	0.01	0.05	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.95	0.99
acceptance	91.4	91.0	91.0	90.0	90.3	89.2	88.5	87.5	86.4	75.3	60.5
blame	91.0	91.4	91.8	91.0	90.3	89.2	88.5	88.2	88.2	78.1	63.4
humor	71.3	72.4	72.0	71.3	69.5	69.9	67.5	67.0	65.2	61.6	57.3
negative	83.8	84.9	86.7	86.7	86.4	85.7	86.0	86.0	85.3	74.9	60.2
positive	89.6	89.6	89.6	88.9	87.5	87.8	87.5	87.5	87.8	76.7	63.8
sadness	59.0	61.6	60.9	61.3	60.6	60.2	58.8	59.5	59.1	57.7	58.5

- In real-world no transcripts, transcription error high so:

#### Decode and classification



#### Lattice classification results

code vs λ	0.01	0.05	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.95	0.99
acceptance	71.4	72.9	75.4	73.6	73.6	73.2	71.8	71.1	68.9	64.6	63.6
blame	75.0	76.8	77.9	78.6	78.2	77.9	76.8	76.4	73.9	67.5	63.9
humor	57.9	58.6	58.6	57.5	57.1	56.4	57.9	56.1	55.4	55.0	50.7
negative	64.3	66.1	69.6	71.1	70.4	69.3	69.3	67.9	65.7	60.7	58.9
positive	72.9	73.2	74.6	74.6	72.5	72.9	73.9	73.6	71.4	66.1	64.6
sadness	52.5	55.0	55.7	52.1	50.4	50.7	51.1	51.8	52.1	54.3	52.1

- Can inform experts:

Word	Most blaming words in terms of discriminative contribution			Least blaming words in terms of discriminative contribution			
	No Bl.	Blame	Δ	Word	No Bl.	Blame	Δ
YOU	-95.49	-85.88	-9.61	EXPECTS	-16.70	-17.84	1.14
YOUR	-51.24	-47.18	-4.06	CONSIDERATION	-16.11	-17.31	1.21
ME	-40.27	-37.74	-2.53	KNOW	-35.10	-36.62	1.53
TELL	-33.97	-32.46	-1.51	INABILITY	-16.76	-18.32	1.55
ACCEPT	-25.44	-23.99	-1.45	SESSION	-20.51	-22.07	1.56
CARING	-27.05	-25.91	-1.14	OF	-44.50	-46.26	1.76
KITCHEN	-21.22	-20.21	-1.02	ANTICIPATION	-22.22	-24.21	2.00
TOLD	-29.04	-28.19	-0.85	THINK	-35.70	-37.77	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

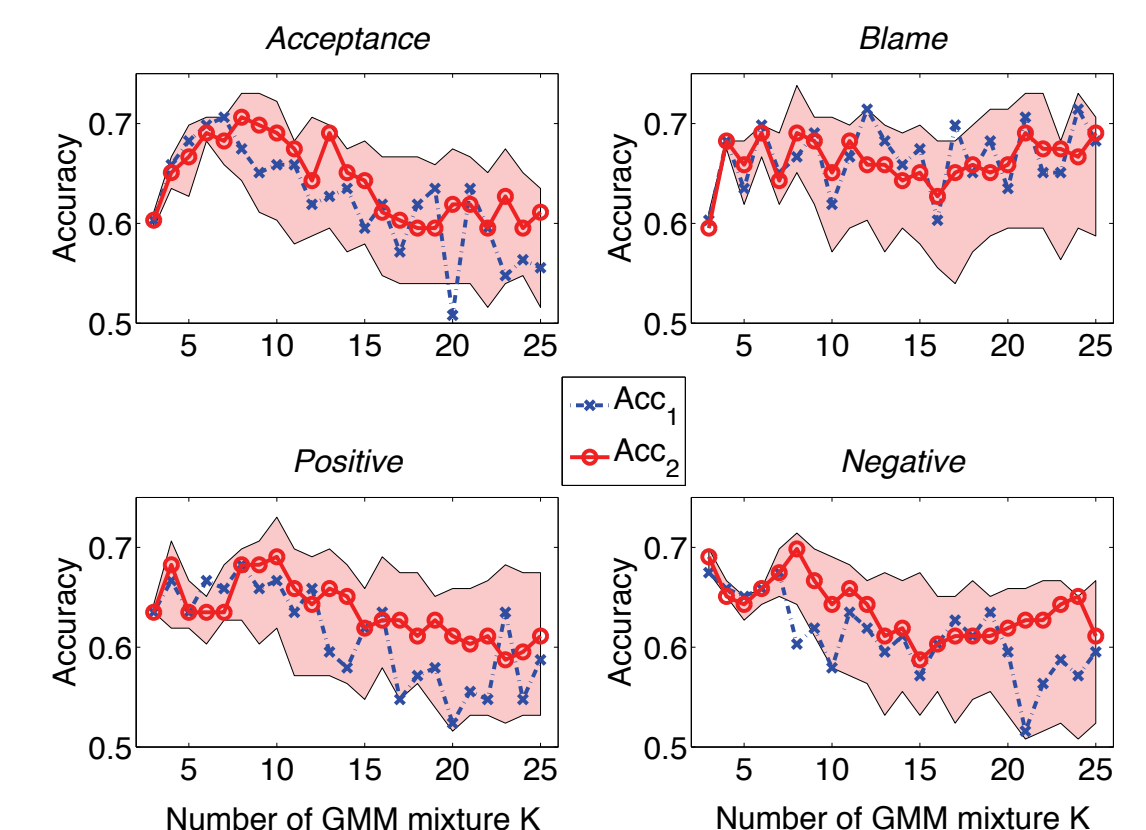
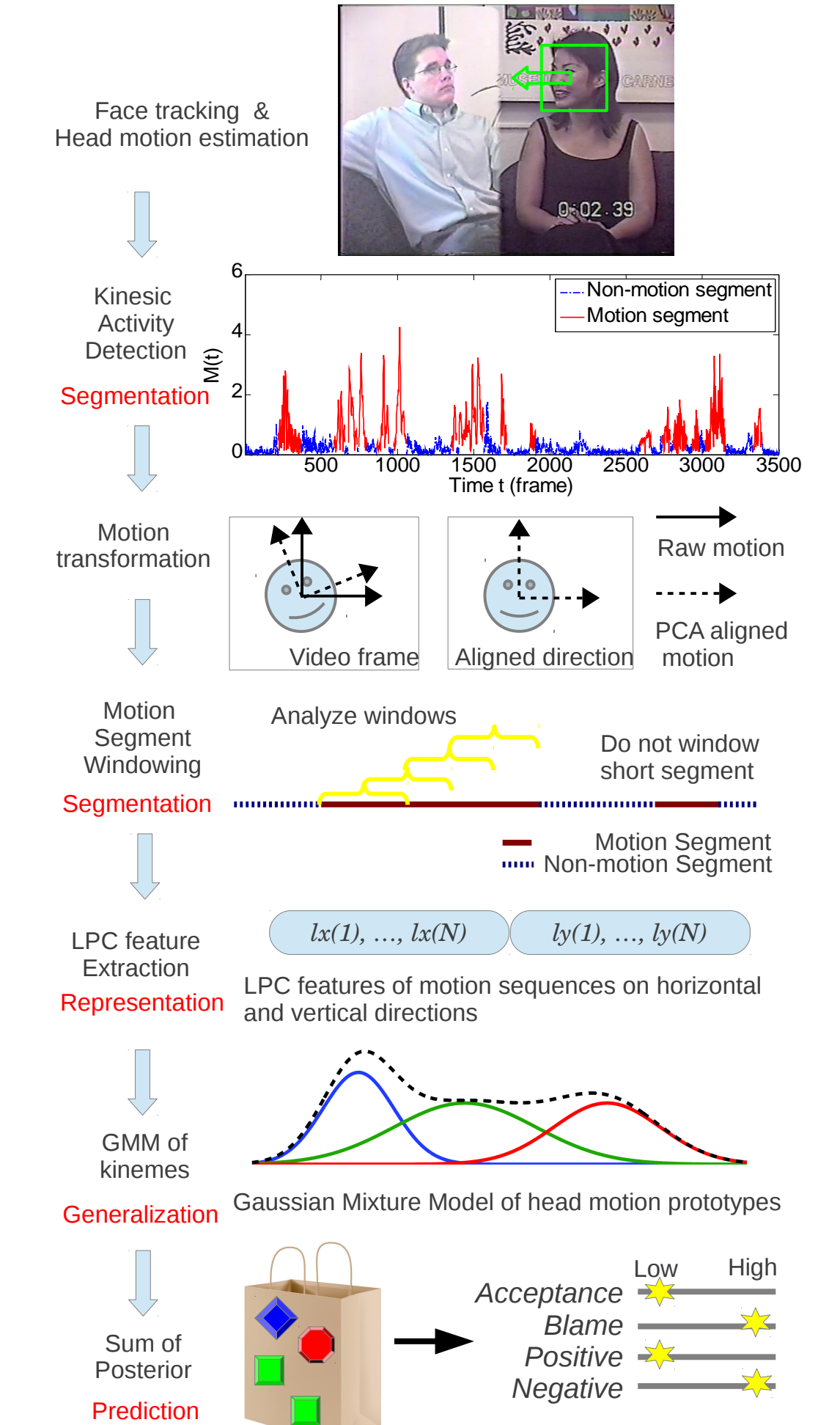
### Future Work Highlights

- Introduce "latent layer" of behavioral primitives
- Improve on individual modalities. e.g. optimize ASR
- Implement fusion based on modality saliencies

### Visual: Head Motion

#### Q: Does head motion capture behavior?

- Continuous, inter- and intra- person variation
- Lacking common list of kinemes
- Implicit link with affect and attitude



### Citations, Acknowledgments

Full list of publications at <http://sail.usc.edu>  
 Work funded by NSF SHB program

