Acoustical Analysis of Engagement Behavior in Children

Rahul Gupta[1], Chi-Chun Lee (Jeremy)[1], Daniel Bone[1], Agata Rozga[2], Sungbok Lee[1], Shrikanth S. Narayanan[1]

Signal Analysis and Interpretation Laboratory (SAIL), University of Southern California[1]
School of Interactive Computing, Georgia Institute of Technology[2]

Motivation & Introduction

- Autism spectrum disorders (ASD) are developmental disorders that result in impaired
  - Social interaction and reciprocity
  - Expressive and receptive language
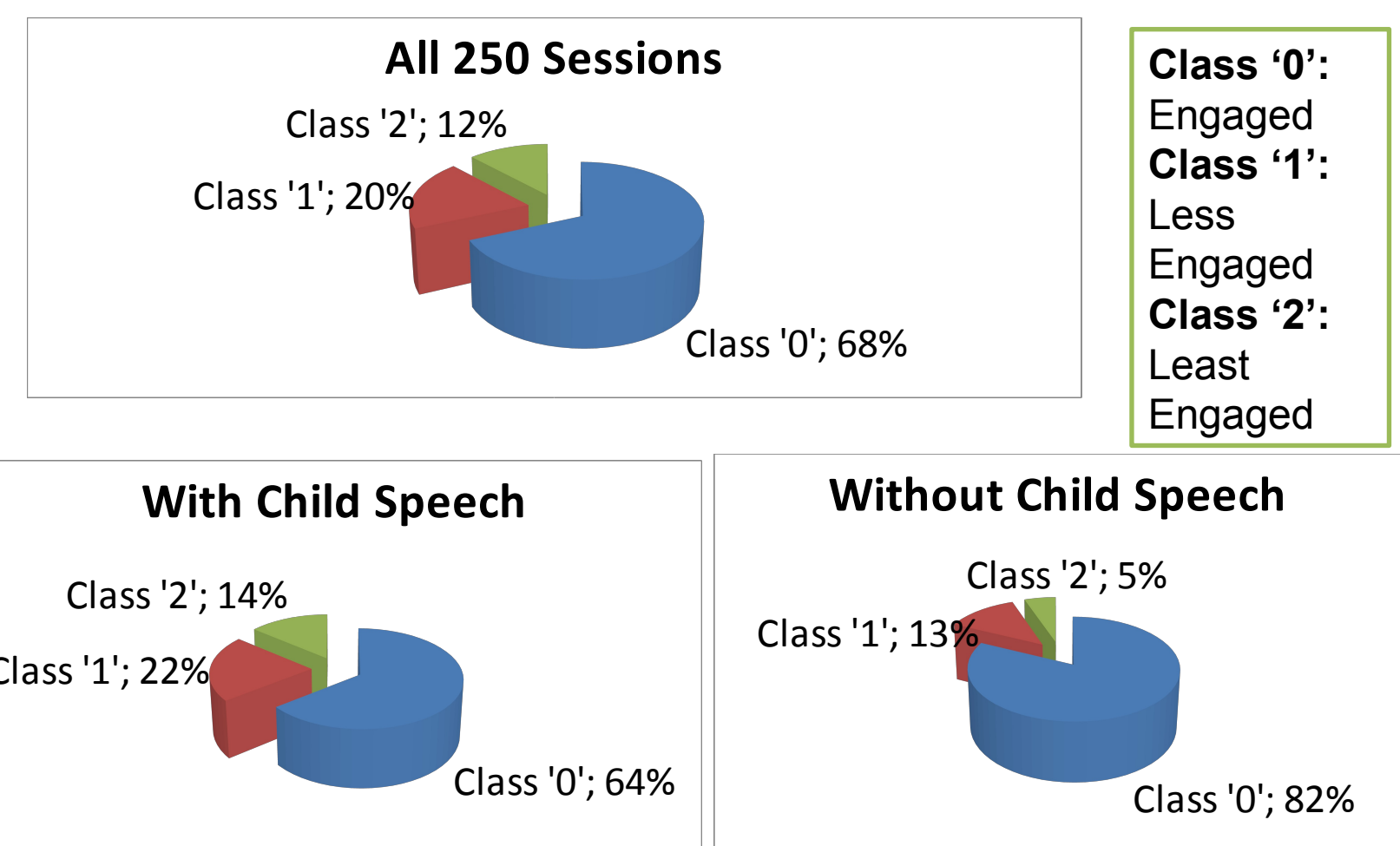 - Restricted and repetitive behavior

- ASD a spectrum disorder due to the *extreme* heterogeneity of symptomatology

- Recent prevalence studies indicate as many as *1 out of 80* children has ASD
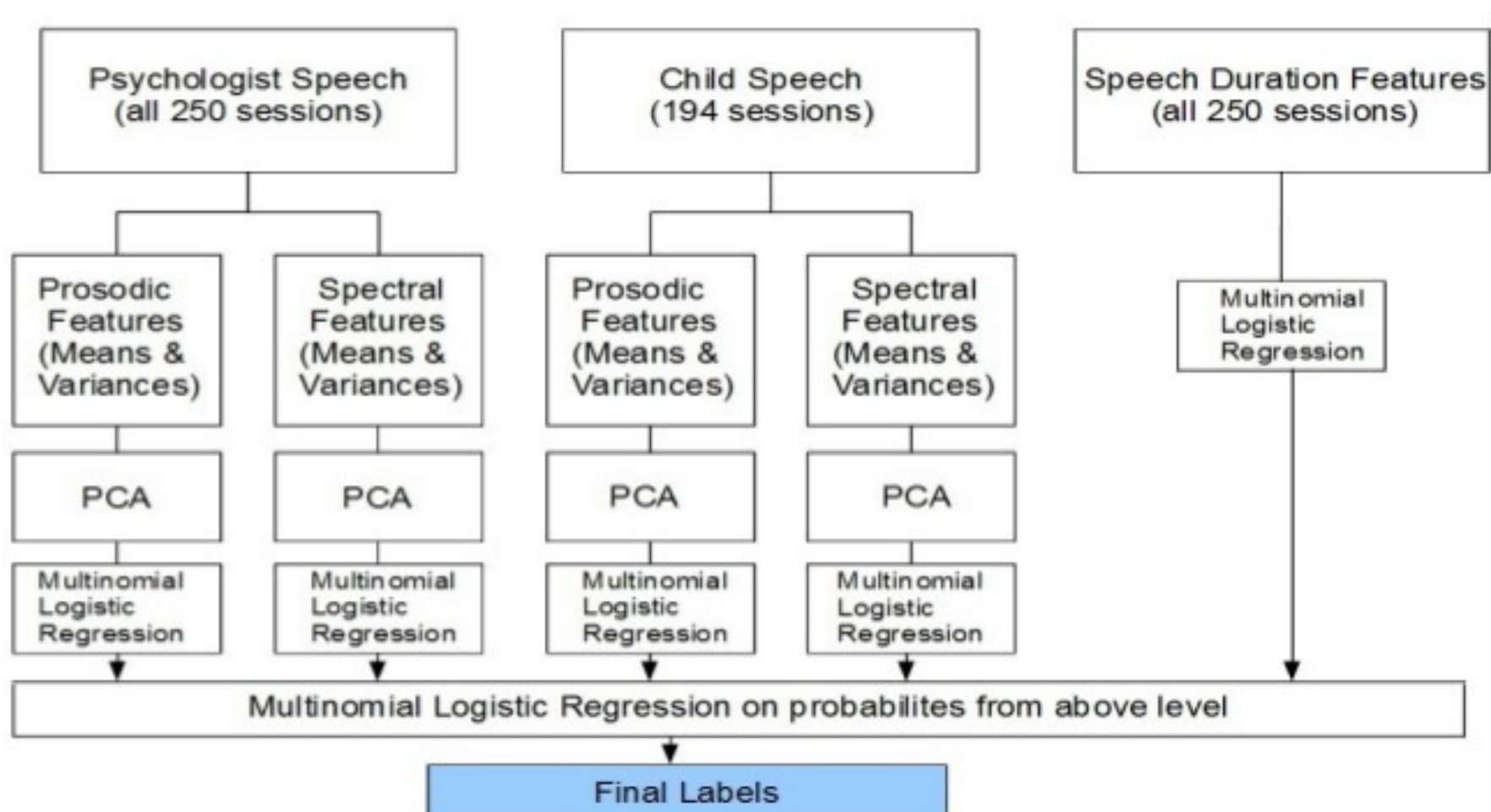
- Joint attention behavior
  - Children's speech and language development [Moore and Dunham, 1995]
  - Characterization of atypical behavior in Autism Spectrum Disorder (ASD) [Mundy et al., 1990]

- Children's engagement behavior
  - Closely related to aspects of joint attention
  - Bring insights into social communicative behavior for child on the spectrum
  - Hypothesis: Engagement level of children is reflected in the vocal cues of the child and the psychologist while participating in tasks largely requiring visual joint attention

The R-ABC database

- Collected as part of a larger NSF funded study
- Non-verbal children (9-30 months old) – Only TD at the moment
- 3-5 minute-long interactive assessment protocol
- 5 different tasks designed to elicit key social communicative behaviors
  - Smiling and saying "hello"
  - Ball play
  - Jointly looking at a book
  - Putting on a book on your head as if it is a hat
  - Smiling and tickling
- 50 sessions of 5 tasks each: 250 sub-sessions
- Audio, video, electro-dermal activity recordings
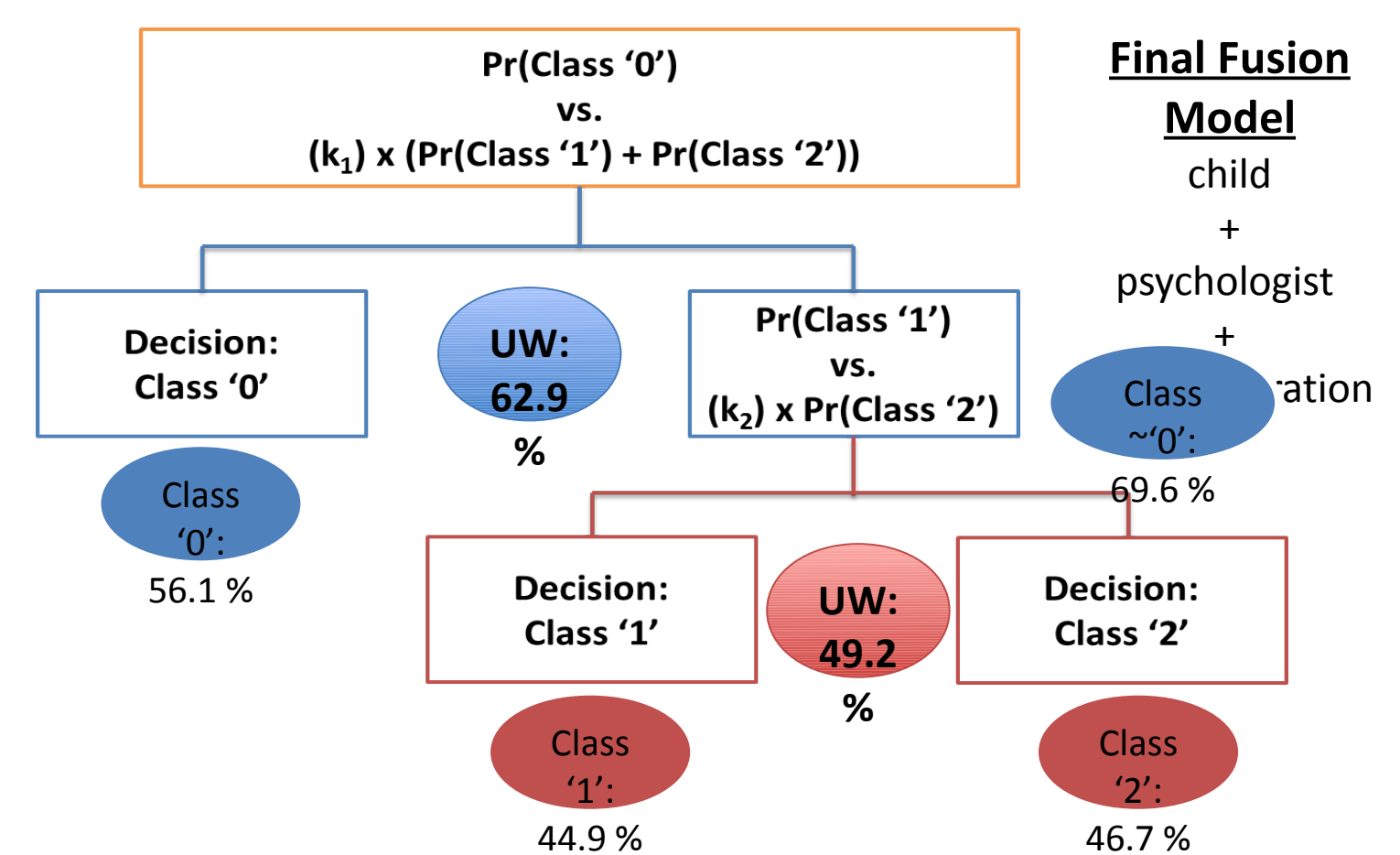- Psychologist rates the child engagement into 3 levels

All 250 Sessions
Class '2'; 12%
Class '1'; 20%
Class '0'; 68%

Class '0': Engaged
Class '1': Less Engaged
Class '2': Least Engaged

With Child Speech
Class '2'; 14%
Class '1'; 22%
Class '0'; 64%

Without Child Speech
Class '2'; 5%
Class '1'; 13%
Class '0'; 82%

Classification Framework

Feature Source	Feature Set	Unweighted Accuracy	Effective Accuracy
Child (194 samples)	Chance	33.3%	
	Spectral	34.1%	Class '0'
	Prosodic	32.6%	Class '1'
	Fused Features	**43.6%**	
Psychologist (250 samples)	Chance	33.3%	
	Spectral	36.3%	Class '0'
	Prosodic	36.4%	Class '0'
	Fused Features	**37.0%**	
Speech Duration		**41.9%**	Class '0' & Class '1'

Class '0': Engaged
Class '1': Less Engaged
Class '2': Least Engaged

Fusion Results

Final Fusion Model
child + psychologist + duration
Class ~'0': 69.6 %

Pr(Class '0') vs. (k1) x (Pr(Class '1') + Pr(Class '2'))

Decision: Class '0' — Class '0': 56.1 %
UW: 62.9 %

Pr(Class '1') vs. (k2) x Pr(Class '2')

Decision: Class '1' — Class '1': 44.9 %
UW: 49.2 %
Decision: Class '2' — Class '2': 46.7 %

Discussion & Future Work

Discussion

- Acoustic features are predictive of child's engagement level
- Easier discrimination between engaged vs. disengaged as compared to two subclasses of disengagement
- Use of other cues (visual, EDA) not included

Future work
- Use of temporal relationship between engagement levels across sub-sessions
- Inclusion of other types of speech cues
- Continuous measure of engagement level

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering