An Overview of Methods in Linear Least-Squares Regression

Sophia Yuditskaya

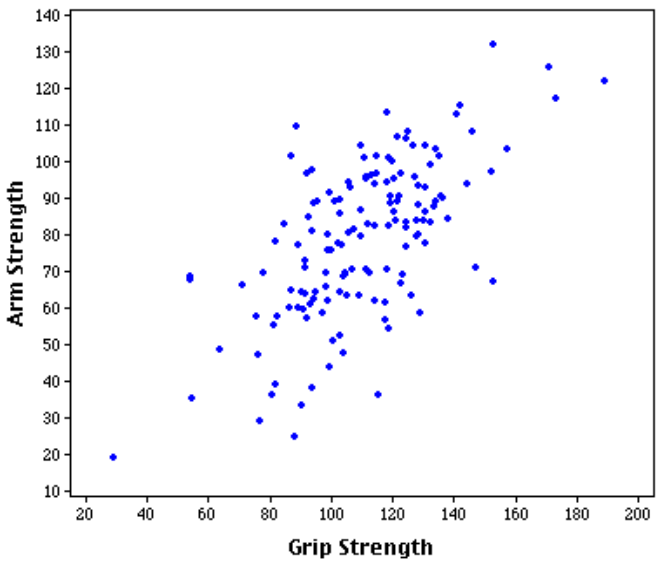
MAS.622J Pattern Recognition and Analysis

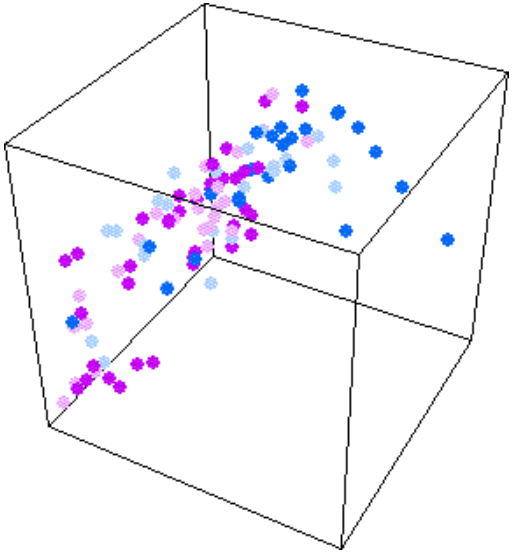
November 4, 2010

Agenda

- Simple Linear Regression
  - Deriving the model
  - Evaluating the model
- Regression with Factor Analysis
  - Principal Components Regression
  - Partial Least Squares Regression
- In-depth Application Example

Data





$$n = \#\,\text{samples}$$

$$k = \#\,\text{independent variables}$$

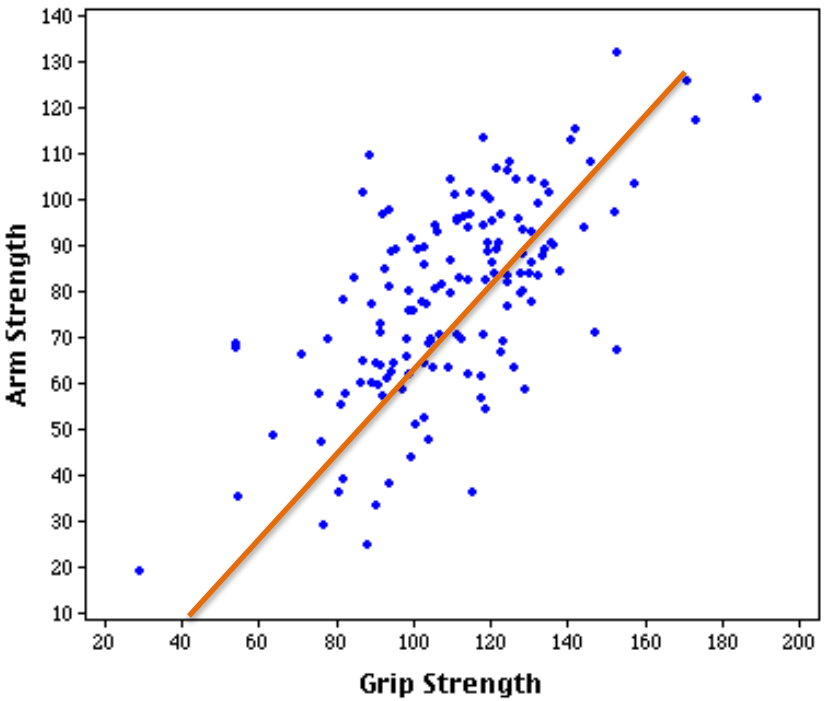
$$m = \#\,\text{dependent variables}$$

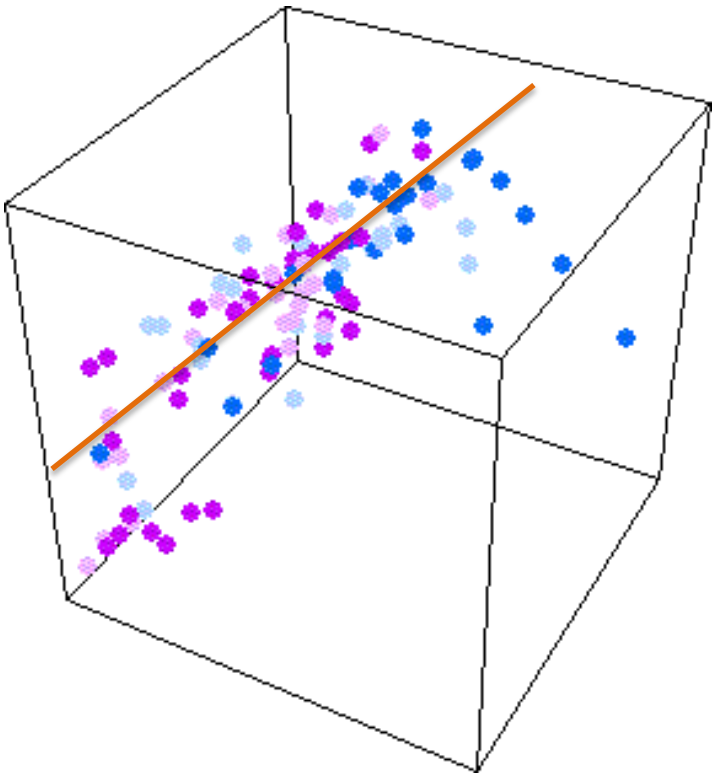
$$X = \begin{bmatrix} predictors \\ n \times k \end{bmatrix}$$

$$Y = \begin{bmatrix} response \\ n \times m \end{bmatrix}$$

Goal: Fit a curve to the data

Linear Regression



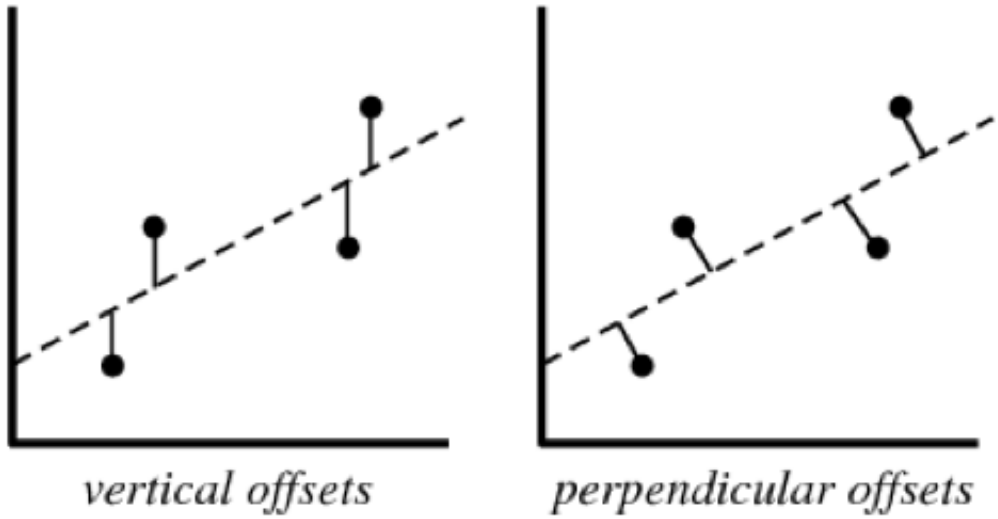


Regression coefficients

$$Y = XB + E$$

Variance unexplained by regression

Offsets (a.k.a. “residuals”)



- Simpler analytic form
- Generalizes better to polynomial fitting

Least-squares Fitting

- Best fit line: $f(x, a_0, a_1) = a_0 + a_1 x$

- Minimizing sum of squares of the vertical offsets:

$$V^2 \equiv \sum_{i=1}^{n} [y_i - f(x_i, a_0, a_1)]^2$$

- Finding the minimum: $\dfrac{\partial (V^2)}{\partial a_k} = 0$

Linear Least-Squares Fitting

$$f(a_0, a_1) = a_0 + a_1 x$$

Solving for a_0 and a_1

$$V^2(a_0, a_1) \equiv \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)]^2$$

$$\frac{\partial(V^2)}{\partial a_0} = -\cancel{2}\sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)] = 0$$

$$\frac{\partial(V^2)}{\partial a_1} = -\cancel{2}\sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)]x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} a_0 x_i - \sum_{i=1}^{n} a_1 x_i^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i$$

$$\sum_{i=1}^{n} y_i = n a_0 + a_1 \sum_{i=1}^{n} x_i$$

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix}$$

Linear Least-Squares Fitting

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \boxed{\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}^{-1}} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \boxed{\begin{bmatrix} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i \\ -\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} y_i x_i \end{bmatrix}}$$

Linear Least-Squares Fitting

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i \\ -\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} y_i x_i \end{bmatrix}$$

$$a_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\left(\sum_{i=1}^{n} y_i x_i\right) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

Alternate Perspective

$$a_1 = \frac{\left(\sum_{i=1}^{n} y_i x_i\right) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)}$$

$$a_0 = \bar{y} - a_1\bar{x}$$

$$V^2(a_0, a_1) \equiv \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)]^2$$

$$\frac{\partial(V^2)}{\partial a_0} = -2\sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)] = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i$$

$$\frac{\sum_{i=1}^{n} y_i}{n} = \frac{n a_0}{n} + \frac{a_1 \sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = a_0 + a_1\bar{x}$$

How Well Does Our Model Fit?

- Proportion of variability in Y that is explained
  - Coefficient of Determination

$$r^2 = 1 - \frac{\sum_{i=1}^{n}\left[y_i - (a_0 + a_1 x_i)\right]^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{\mathrm{cov}(x, y)^2}{\mathrm{var}(x)\,\mathrm{var}(y)}$$

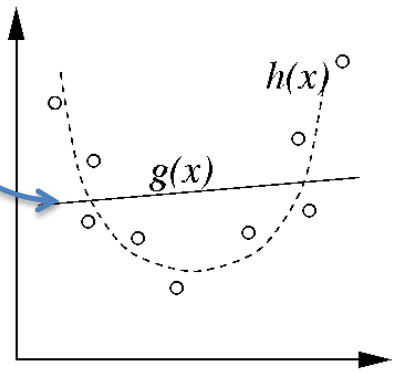
  - Correlation Coefficient

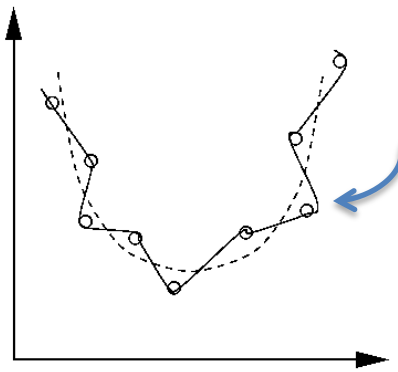
$$r = \sqrt{r^2} = \sqrt{\frac{\mathrm{cov}(x, y)^2}{\mathrm{var}(x)\,\mathrm{var}(y)}} = \frac{\mathrm{cov}(x, y)}{stdev(x)\,stdev(y)}$$

Problem of Overfitting

- Fitting true patterns or noise?

Underfitting





Overfitting

Predictors/Features

+

Samples

Overfitting

Real-World Data

- Constrained sample size
  - Data collection can be challenging, expensive
    - e.g. Manual human annotation
  - Outliers, inconsistencies, annotator disagreements
    - Unreliable for modeling, thrown out
- Can be highly multidimensional
  - Lots of features (i.e. predictor variables)
- Multicollinearity
  - Some predictors are highly correlated with others
  - Creates noise in the data

Solution?

- Reduce dimensionality
- Eliminate multicollinearity
  - Produce a set orthogonal predictor variables

Principal Component Analysis (PCA)

Factor Analysis

Factor Analysis

Given predictors X, response Y

Goal: Derive a model B such that Y=XB+E, with factor analysis to reduce dimensionality

1. Compute new factor (i.e. "component") space W
2. Project X onto this new space to get *factor scores*:

$$T = XW$$

3. Represent the model as:

$$Y = TQ+E = XWQ+E$$

...where B = WQ

Principal Components Regression (PCR)

- Applies PCA to predictors prior to regression
- Pros:
  - Eliminates collinearity → orthogonal space
  - Produces model that generalizes better than MLR
- Cons:
  - May be removing signal & retaining noise
  - How to retain the predictors X that are most meaningful in representing response Y?

Partial Least Squares Regression (PLS)

- Takes into account Y in addition to X
- A different kind of Factor Analysis

  Recall, T=XW…

  - PCR: W reflects var(X)
  - PLS: W reflects cov(X, Y)
- Multiple Linear Regression (MLR) vs. PCR vs. PLS
  - MLR maximizes correlation between X & Y
  - PCR captures maximum variance in X
  - PLS strives to do both by maximizing cov(X, Y)

PLS Factor Analysis

- Two iterative methods
  - NIPALS → Nonlinear Iterative Partial Least Squares
    - Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.). Multivariate Analysis. (pp.391-420) New York: Academic Press.
  - SIMPLS → more efficient, optimal result
    - Supports multivariate Y
    - De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18: 251–263

SIMPLS

Inputs

$s = \#\,\text{samples}$

$m = \#\,\text{predictor variables}$

$n = \#\,\text{response variables}$

$c = \#\,\text{components to factor}$

$$X = \begin{bmatrix} predictors \\ s \times m \end{bmatrix} \qquad Y = \begin{bmatrix} response \\ s \times n \end{bmatrix}$$

Outputs to be computed

$$W = \begin{bmatrix} empty \\ m \times c \end{bmatrix} = \text{Weight matrix for X, such that } T = XW$$

$$T = \begin{bmatrix} empty \\ s \times c \end{bmatrix} = \text{PLS factor scores matrix}$$

$$P = \begin{bmatrix} empty \\ m \times c \end{bmatrix} = \text{Factor loadings matrix for X, such that } X = TP^T + F$$

$$Q = \begin{bmatrix} empty \\ n \times c \end{bmatrix} = \text{Factor loadings matrix for Y, such that } Y = TQ^T + E$$

$$B = \begin{bmatrix} empty \\ m \times n \end{bmatrix} = \text{PLS regression coefficients of Y on X}$$

Unexplained variance in X

Unexplained variance in Y

c = # components to factor

T = [*empty*] = PLS factor scores matrix

W = [*empty*] = Weight matrix for X such that $T = XW$

P = [*empty*] = Factor loadings matrix for X, such that $X = TP + F$

Q = [*empty*] = Factor loadings matrix for Y, such that $Y = TQ + E$

B = [*empty*] = PLS regression coefficients of Y on X

$$A_0 = X^T Y$$
$$M_0 = X^T X$$
$$C_0 = I$$

for $h = 1, \ldots, c$

$q_h = \text{dominant eigenvector of } A_{h-1}^T A_{h-1}$

$w_h = A_{h-1} q_h$

$c_h = w_h^T M_{h-1} w_h$

$w_h = \dfrac{w_h}{\sqrt{c_h}} \rightarrow$ store into column h of W

$p_h = M_{h-1} w_h \rightarrow$ store into column h of P

$q_h = A_{h-1}^T w_h \rightarrow$ store into column h of Q

$v_h = C_h p_h$

$v_h = \dfrac{v_h}{\|v_h\|}$ (normalize)

$C_h = C_{h-1} - v_h v_h^T$

$M_h = M_{h-1} - p_h p_h^T$

$A_h = C_h A_{h-1}$

end loop

Finally…

P, Q, W already assembled, iteratively

Compute $T = XW$

Compute $B = WQ^T$

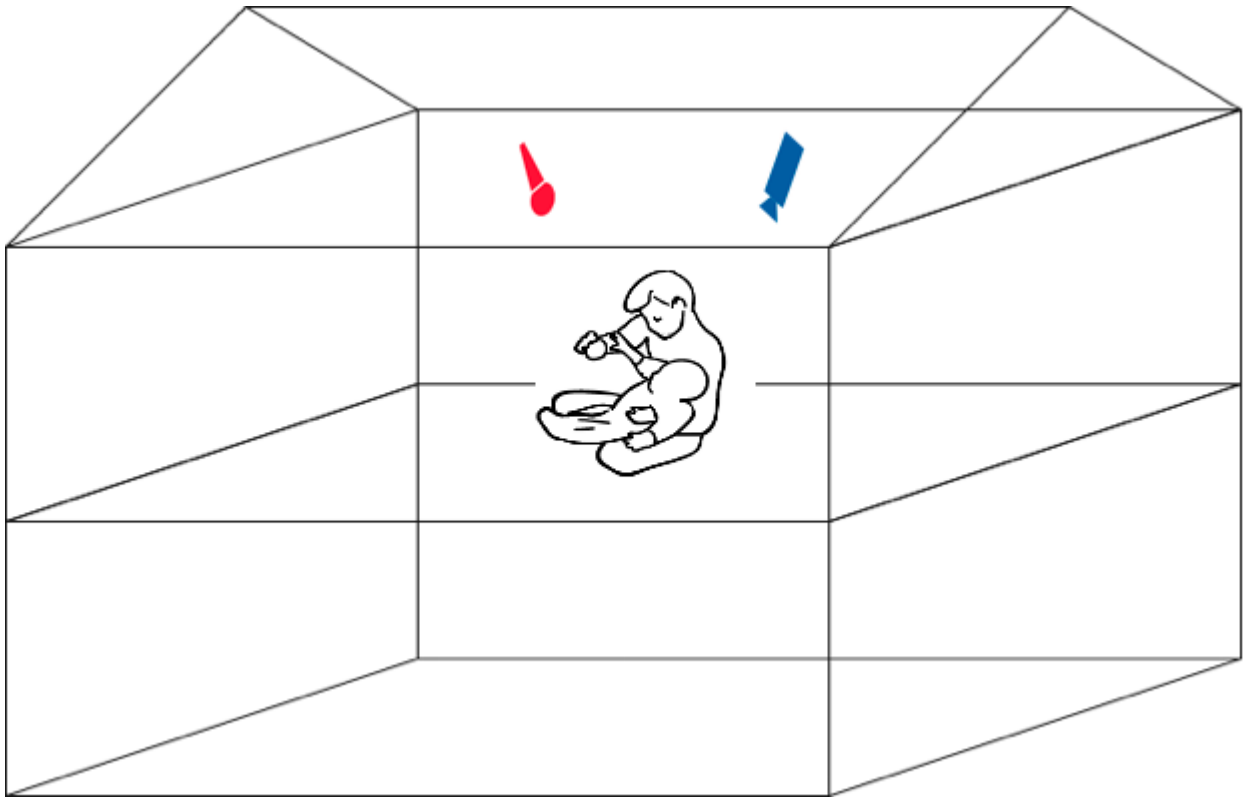
Application Example

Automatic Vocal Recognition of a Child's Perceived Emotional State within the Speechome Corpus

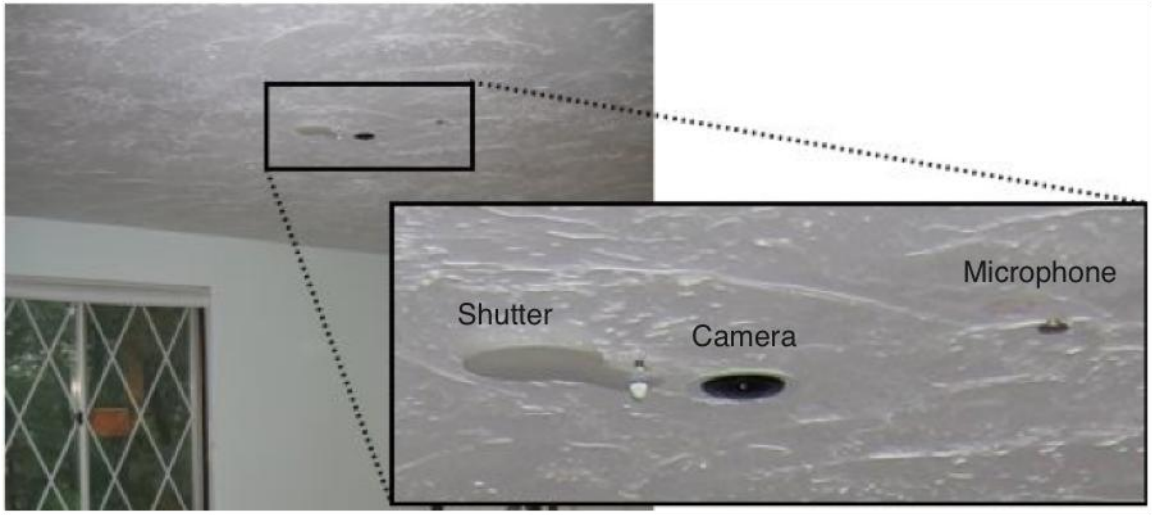
Sophia Yuditskaya

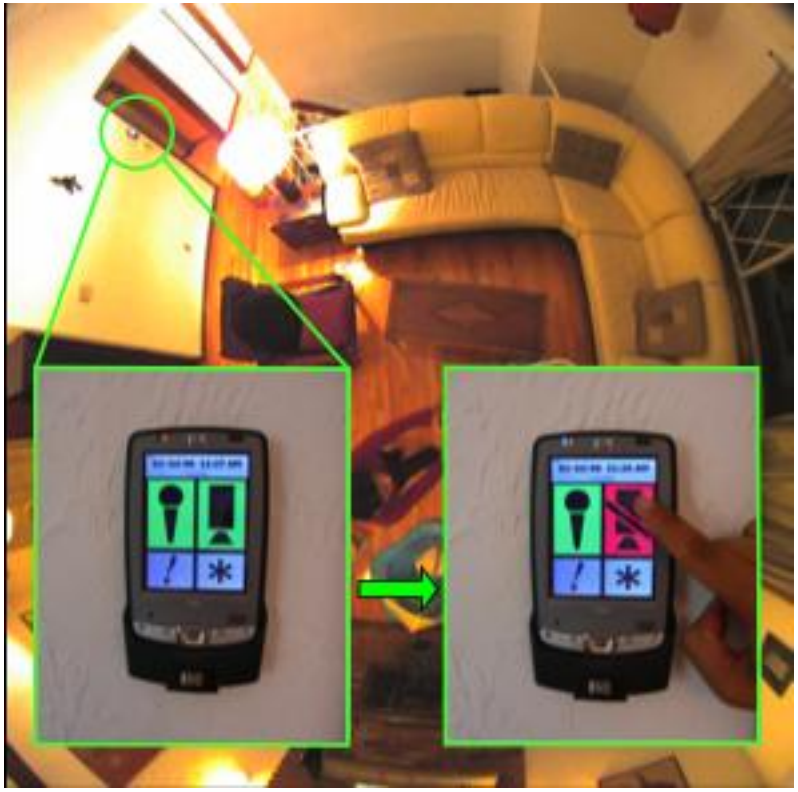
M.S. Thesis (August 2010)
MIT Media Laboratory

The Human Speechome Project



The Human Speechome Project





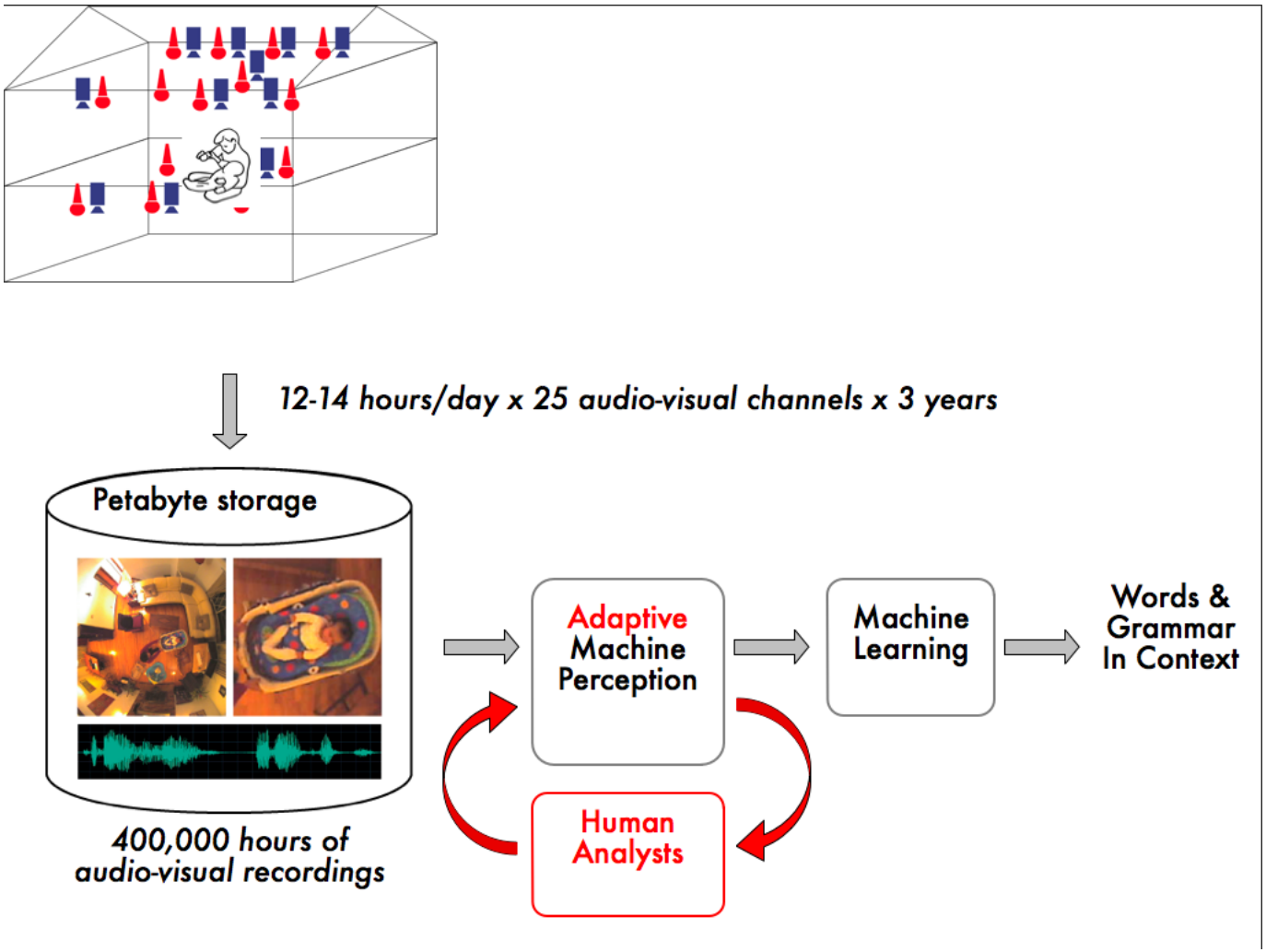
Recording Statistics

- Birth to age three
- ~ 9.6 hours per day
- 70-80% of child’s waking life
- Over 230,000 hours of A/V recordings

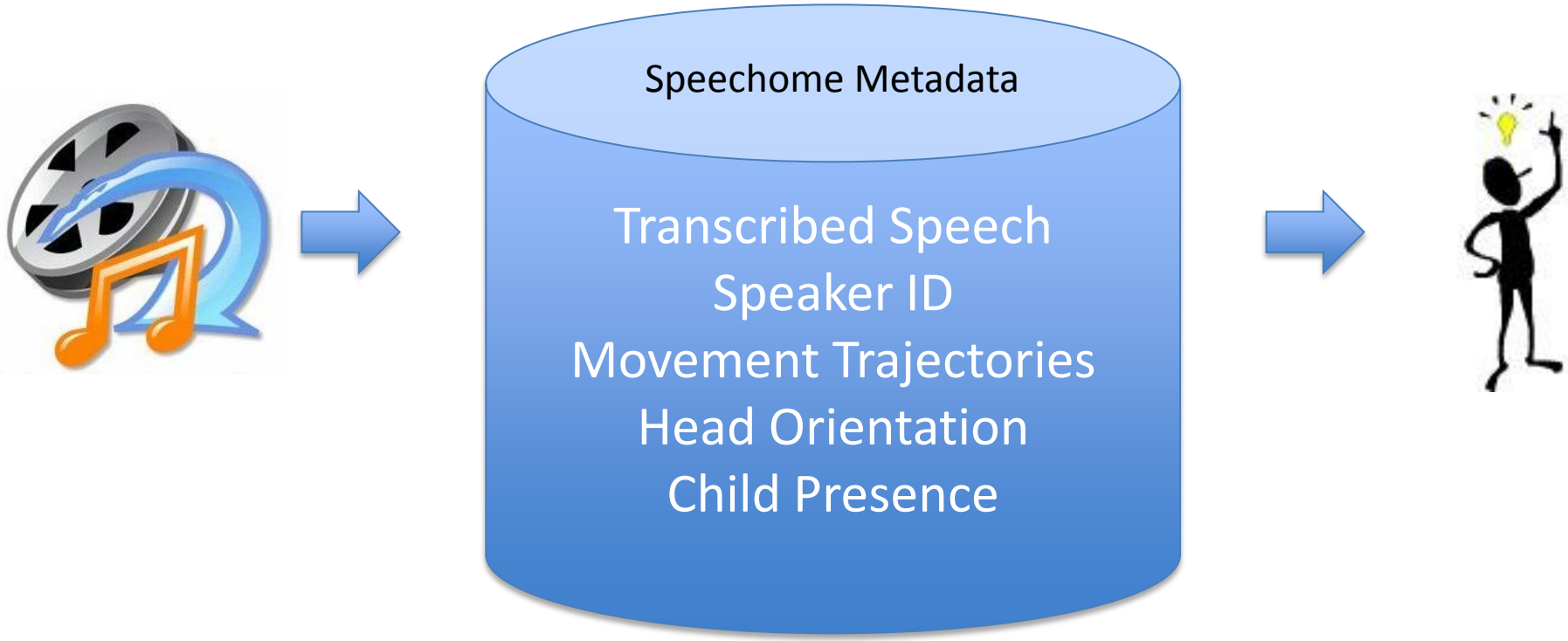
Speechome:
A New Kind of Behavioral Data

- Ecologically Valid
  - Familiar home environment
  - Child-caretaker interaction
  - Real daily life situations
- Ultra-dense, longitudinal
  - Comprehensive record
  - Insights at variety of time scales
    - (e.g.) Hourly, Daily, Weekly, Monthly, Yearly
    - Up to millisecond precision

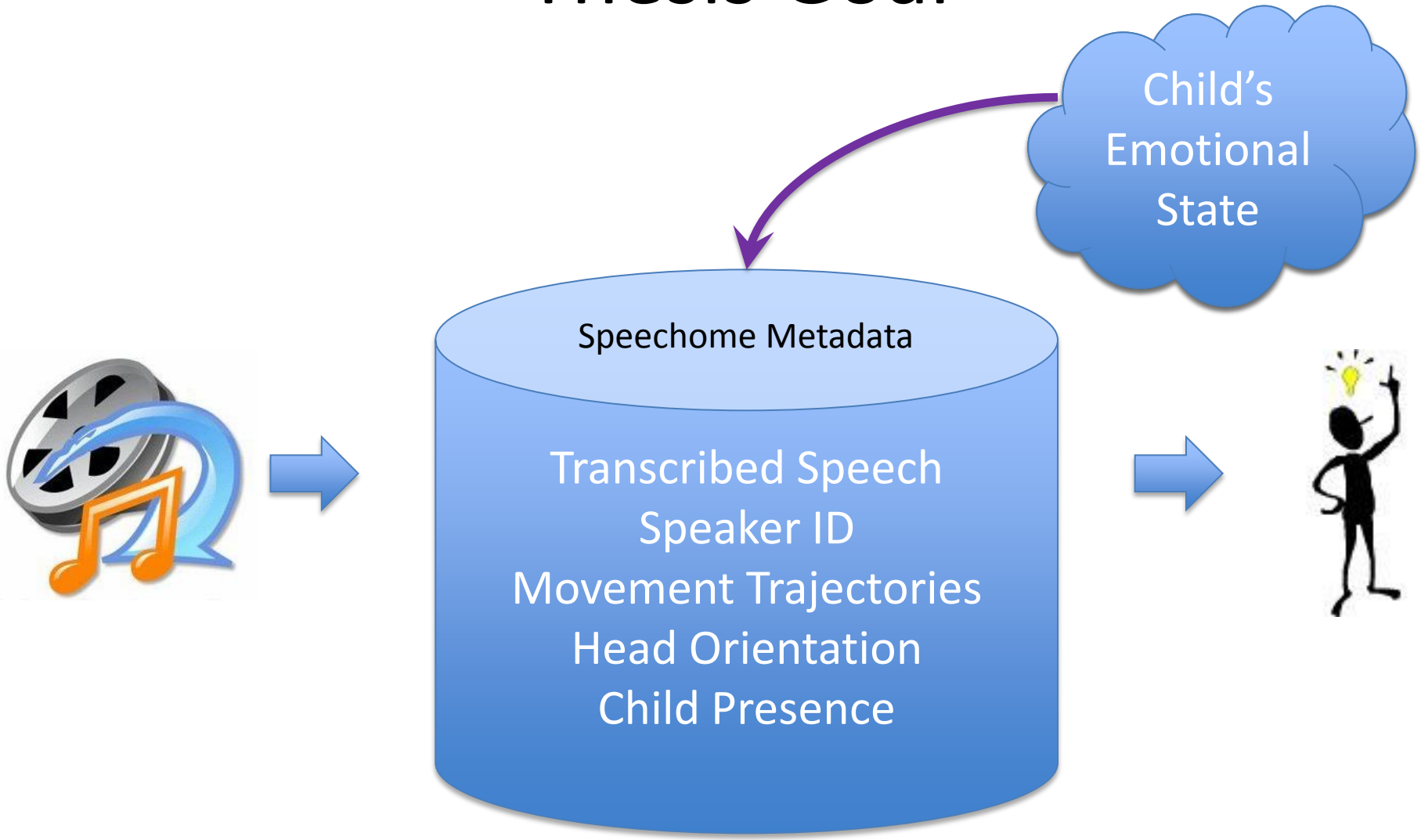
Challenges


12-14 hours/day x 25 audio-visual channels x 3 years
Petabyte storage
400,000 hours of audio-visual recordings
Adaptive Machine Perception
Human Analysts
Machine Learning
Words & Grammar In Context

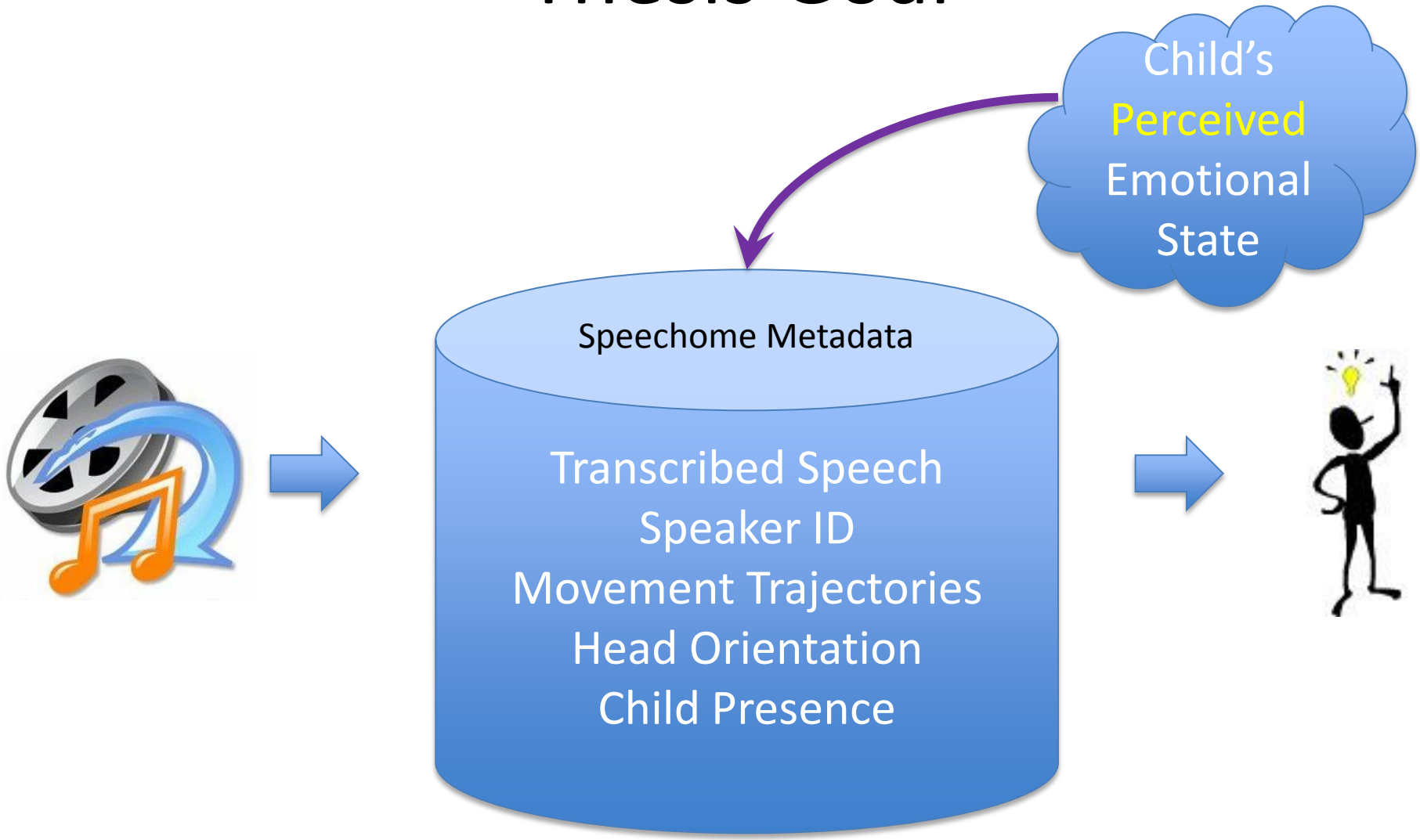
Previous & Ongoing Efforts



Thesis Goal


Child's Emotional State
Speechome Metadata
Transcribed Speech
Speaker ID
Movement Trajectories
Head Orientation
Child Presence

Thesis Goal

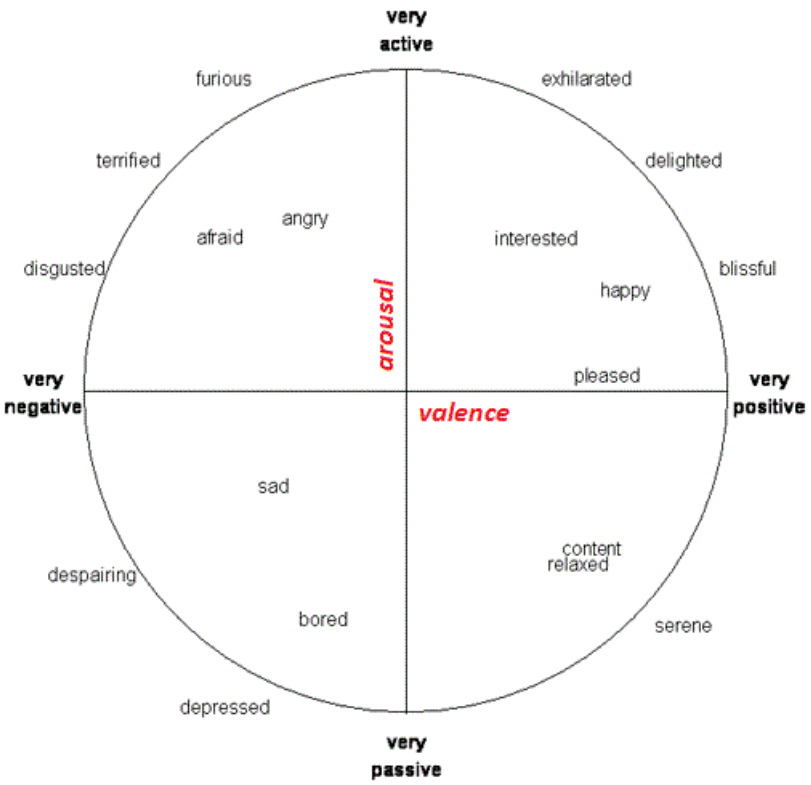

Child's
Perceived
Emotional
State
Speechome Metadata
Transcribed Speech
Speaker ID
Movement Trajectories
Head Orientation
Child Presence

Questions for Analysis

- Developing a Methodology
  - How to go about creating a model for automatic vocal recognition of a child's perceived emotional state?
- Evaluation
  - How well can our model simulate human perception?
- Exploration
  - Socio-behavioral
    - Do correlations change in **social situations** or when the child is **crying**, **babbling**, or **laughing**?
  - Dyadic
    - How much does **proximal adult speech** reflect the child's emotional state?
    - Do certain **caretakers** correlate more than others?
  - Developmental
    - Are there any **developmental trends** in correlation?

Methodology

Representing Emotional State



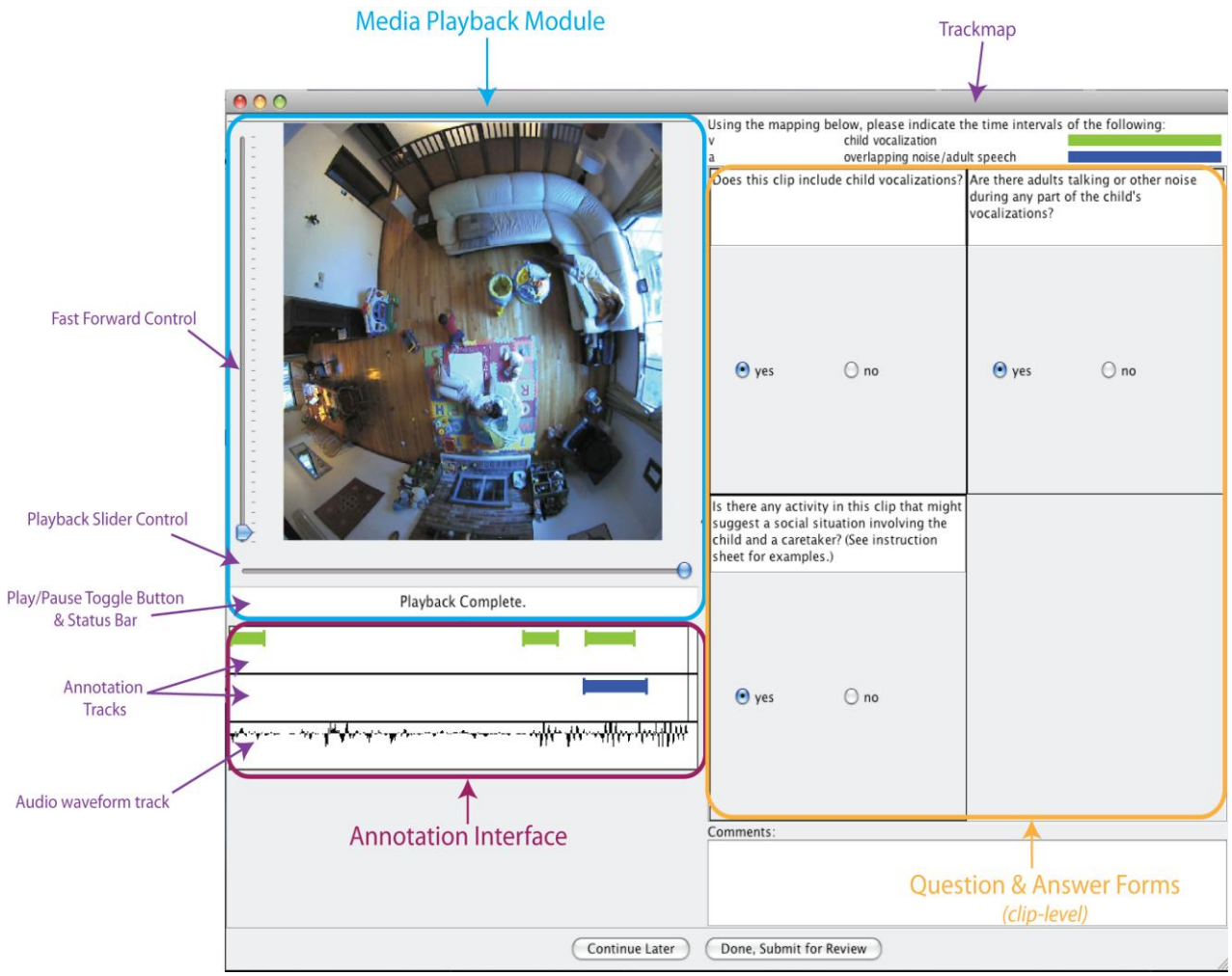
Emotional State = <valence (mood), arousal (energy)>

Ground Truth Data Collection

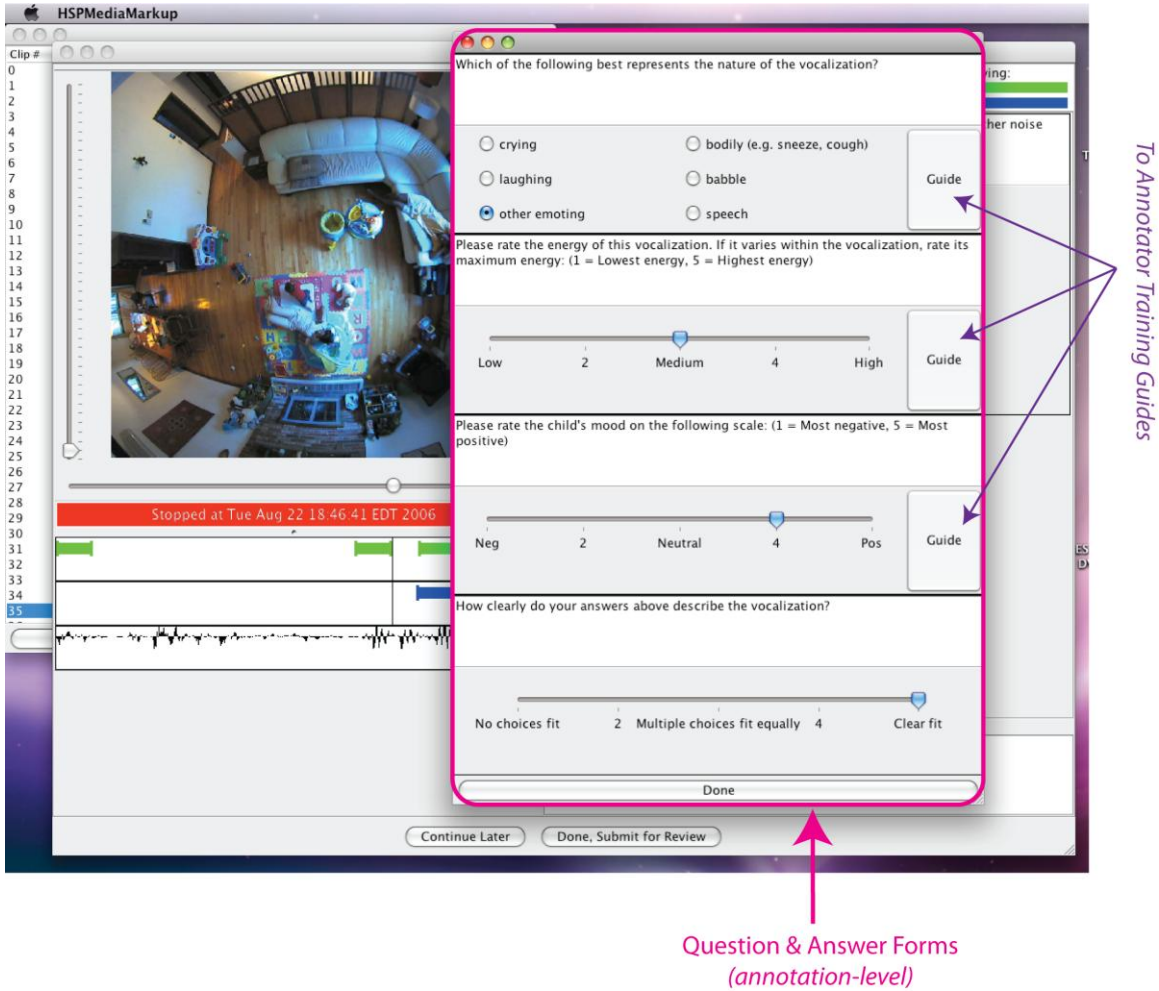
Data Collection Methodology

- Annotation interface
  - Modular, configurable
  - Streamlined for efficient annotation
  - Direct streaming Speechome A/V
- Supporting infrastructure
  - Administrative tool
  - Database
- Questionnaire
- Data mining strategies to minimize annotation volume
- Hired & supervised annotators (UROPs)

Annotation Interface



Annotation Interface



Minimizing Annotation Volume

- Longitudinal Sampling strategy
  - 2 full days / mo: months 9-15
  - 1 full day / 3 mo: months 15-24
- Filter using existing metadata
  - Detected speech + Child presence
  - Speaker ID (machine learning solution)

Speaker ID

<Speaker Label, Confidence Value>

5 Speaker Categories:

- Child
- Mother
- Father
- Nanny
- Other – e.g. Toy sounds, sirens, speakerphone

Filtering by Speaker ID

<Speaker Label, Confidence Value>

5 Speaker Categories:

- Child → Keep all
- Mother
- Father
- Nanny
- Other – e.g. Toy sounds, sirens outside, speakerphone

Apply Filtering

Problem: choose appropriate confidence thresholds

-Above the threshold → Filter out

-Below the threshold → Keep

Filtering by Speaker ID

1-Specificity

Confidence Threshold Configuration				# Segments After Filtering		Filtering Accuracy						
Father	Nanny	Mother	Other	Accepted	Rejected	TP	FP	TN	FN	$\frac{TP}{(TP+FN)}$	$\frac{FP}{(FP+TN)}$	Filter Ratio
0.60	0.75	0.45	0.55	5500	2356	2496	3004	2121	235	0.914	0.586	0.30
0.55	0.75	0.35	0.65	5107	2749	2412	2695	2430	319	0.883	0.525	0.35
0.70	0.50	0.40	0.45	4714	3142	2327	2387	2738	404	0.852	0.466	0.40
0.35	0.70	0.35	0.50	4321	3535	2239	2082	3043	492	0.820	0.406	0.45
0.55	0.45	0.25	0.50	3929	3927	2127	1802	3323	604	0.779	0.351	0.50
0.35	0.40	0.45	0.40	3536	4320	2015	1521	3604	716	0.738	0.297	0.55

$$sensitivity = \frac{TP}{(TP + FN)} \qquad specificity = \frac{TN}{(FP + TN)}$$

Criteria: (in order of priority)

1. Maximize Sensitivity: % child vocalizations retained
2. Minimize (1- Specificity): % irrelevant clips retained

Filtering by Speaker ID

- 22% reduction in volume
- Saved roughly 87 hours of annotation time

Date	Total # Segments	Accepts	Rejects	Filtering Ratio
05/16/2006	1807	1336	471	0.260
05/21/2006	1064	750	314	0.295
06/11/2006	1361	919	442	0.325
06/24/2006	791	554	237	0.300
07/02/2006	1220	845	376	0.308
07/10/2006	1619	1103	516	0.319
08/07/2006	1323	1003	319	0.241
08/22/2006	2014	1607	407	0.202
09/06/2006	1118	807	311	0.278
09/17/2006	2130	1468	662	0.311
10/05/2006	1982	1622	360	0.182
10/19/2006	1884	1554	330	0.175
01/04/2007	1755	1423	332	0.189
04/09/2007	3227	2471	756	0.234
04/27/2007	2689	2121	568	0.211
07/06/2007	2180	1869	311	0.143

Table 3-2. Applied Speaker ID Filtering Results. Highlighted in blue are the days for which the chosen confidence threshold configuration was applied to filter out irrelevant segments from the annotation dataset. Filtering out 100 segments saves roughly 2 hours of annotation time. Overall, filtering by speaker ID reduced the amount of segments to be heard by annotators by 22%, for a total time-savings of roughly 87 hours.

Annotators

ID	Gender	Class Year	School	Major	Started
1	Female	Freshman	MIT	Math	1/10
2	Female	Freshman	MIT	Biology	1/10
3	Male	Freshman	MIT	EECS	2/10
4	Male	Sophomore	MIT	Math/EECS	1/10
5	Female	Sophomore	MIT	EECS	1/10
6	Female	Junior	MIT	Brain & Cognitive Science	1/10
7	Female	Senior	Wellesley	Neuroscience	2/10

Table 3-3. Annotator Demographics.

Inter-Annotator Agreement

- P(Agreement)
  - Does not account for random agreement
- P(Error)
  - Probability of two annotators agreeing by chance
- Cohen's Kappa

$$\kappa = \frac{P(Agreement) - P(Error)}{1 - P(Error)}$$

Poor agreement = Less than 0.20
Fair agreement = 0.20 to 0.40
Moderate agreement = 0.40 to 0.60
Good agreement = 0.60 to 0.80
Very good agreement = 0.80 to 1.00

Questions

- What is the **Nature** of the vocalization?
  - » Crying, Laughing, Bodily, Babble, Speech, Other Emoting
- Does it occur during a **Social** situation?
- Rate its **Energy** from 1 to 5
  - » 1 = Lowest energy, 3 = Medium, 5 = Highest Energy
- Rate the child's **Mood** from 1 to 5
  - » 1 = Most negative, 3 = Neutral, 5 = Most positive

Rating Scale Variants

- Original 5-point scale: $\{1, 2, 3, 4, 5\}^{5pt}$
- Collapsed 3-point scale: $\{1, 2, 3\}^{3pt}$
  - Version A

    $\{1,2\}^{5pt} \rightarrow \{1\}^{3pt}$, $\{3\}^{5pt} \rightarrow \{2\}^{3pt}$, and $\{4,5\}^{5pt} \rightarrow \{3\}^{3pt}$

  - Version B

    $\{1\}^{5pt} \rightarrow \{1\}^{3pt}$, $\{2,3,4\}^{5pt} \rightarrow \{2\}^{3pt}$, and $\{5\}^{5pt} \rightarrow \{3\}^{3pt}$

- Expanded 9-point scale: $\{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}^{9pt}$
  - $\{$ Disagreed by 1 point $\}^{5pt} \rightarrow \{$ Agreed with average value $\}^{9pt}$

Agreement Analysis

Poor agreement = Less than 0.20
Fair agreement = 0.20 to 0.40
Moderate agreement = 0.40 to 0.60
Good agreement = 0.60 to 0.80
Very good agreement = 0.80 to 1.00

Label	Total # Child Vocalizations	# Agreed	P(Agreement)	P(Error)	Cohen's Kappa (κ)
Social	11518	9518	0.826	0.612	0.552
Nature	11470	7972	0.695	0.391	0.499
Energy (5 pt scale)	11543	5537	0.480	0.253	0.304
Energy (3 pt scale A)	11543	6996	0.576	0.317	0.422
Energy (3 pt scale B)	11543	9470	0.820	0.712	0.376
Energy (9 pt scale)	11543	10543	0.913	0.142	0.899
Mood (5 pt scale)	11544	5838	0.506	0.273	0.321
Mood (3 pt scale A)	11544	6856	0.593	0.306	0.413
Mood (3 pt scale B)	11544	10298	0.892	0.840	0.324
Mood (9 pt scale)	11544	11000	0.953	0.177	0.943

Table 3-5. Agreement Calculations. Very good agreement is highlighted in yellow (9 point scale for both Energy and Mood). Moderate agreement is highlighted in green.

Post-processing

- Generated agreement indexes
- Pruned noise & overlapping adult speech
- Extracted audio clips from Speechome corpus
  - Pruned child vocalizations
  - Adult speech surrounding each voc
    - 30 sec window before, after

Feature Extraction

Features

- Computed 82 acoustic features
- Praat

Attribute	Metrics	Units	Interpolation
Intensity	Min, Mean, Max, Stdev	dB	Parabolic
Pitch	Min, Mean, Max, Stdev	Hertz	Parabolic
	Mean absolute slope	Semitones	n/a
Stylized Pitch Contour	# stylized pitch points	Quantity	n/a
Fast Fourier Transform (FFT)	Centroid, Stdev	Hertz	n/a
	Skewness, Kurtosis	Unitless	
Long-Term Average Spectrum (LTAS)	Min, Mean, Max, Stdev	dB	None
	Frequency of Min, Frequency of Max	Hertz	None
Harmonics-to-Noise Ratio (HNR)	Min, Mean, Max, Stdev	dB	Parabolic
Harmonics-to-Noise Ratio (HNR)	Time of Min, Time of Max	Seconds	Parabolic
Mel-Frequency Cepstral Coefficients (MFCCs), #1 – 16	Mean, Stdev (separately for each MFCC)	Mel	n/a
Formants, #1 – 5	Min, Mean, Max, Stdev	Hertz	Parabolic

Table 4-1. Acoustic Features Extracted for Analysis

Final Dataset for Analysis

- 112 subsets x 5 feature sets = 560 models

Socio-Behavioral Context	Sample Size										
	Total	May 06	Jun 06	Jul 06	Aug 06	Sep 06	Oct 06	Jan 07	Apr 9 07	Apr 27 07	Jul 07
All Vocalizations	7376	527	495	911	965	741	843	625	1175	582	512
Social Situations	3857	284	187	522	501	337	562	367	638	251	208
All Nonbodily	4986	428	350	762	708	513	609	405	665	294	252
Social Nonbodily	3796	258	187	508	498	334	556	365	634	248	208
All Crying	398	70	22	109	48	59	25	23	27	3	12
Social Crying	336	41	17	103	47	40	25	21	27	3	12
All Laughing	57	4	3	10	8	8	2	2	7	11	2
Social Laughing	52	4	2	7	7	8	2	2	7	11	2
All Babble	656	4	19	26	103	33	67	53	144	120	87
Social Babble	506	3	8	17	63	23	59	47	130	96	60
All Speech	442	0	0	0	0	1	0	73	245	53	70
Social Speech	418	0	0	0	0	1	0	54	241	53	69
All Other Emoting	3433	350	306	617	549	412	515	254	242	107	81
Social Other Emoting	2484	210	160	381	381	262	470	241	229	85	65

Table 4-3. Sample sizes, per situational and monthly subsets of the dataset. Subsets highlighted in yellow were retained for analysis.

For each sample, feature sets were computed from:

(1) The child vocalization

(2) Adult speech ***before*** the vocalization

(3) Adult speech ***after*** "

(4) Adult speech ***surrounding*** "

(5) All of the above combined

Building the Model

What is Best for Our Data?

- Highly multivariate input
  - 82 features, 328 combined
- Multivariate output: <mood, energy>
- Small and widely varying sample sizes

Dangers: Overfitting, Statistical Shrinkage

Socio-Behavioral Context	Sample Size										
	Total	May 06	Jun 06	Jul 06	Aug 06	Sep 06	Oct 06	Jan 07	Apr 9 07	Apr 27 07	Jul 07
All Vocalizations	7376	527	495	911	965	741	843	625	1175	582	512
Social Situations	3857	284	187	522	501	337	562	367	638	251	208
All Nonbodily	4986	428	350	762	708	513	609	405	665	294	252
Social Nonbodily	3796	258	187	508	498	334	556	365	634	248	208
All Crying	398	70	22	109	48	59	25	23	27	3	12
Social Crying	336	41	17	103	47	40	25	21	27	3	12
All Laughing	57	4	3	10	8	8	2	2	7	11	2
Social Laughing	52	4	2	7	7	8	2	2	7	11	2
All Babble	656	4	19	26	103	33	67	53	144	120	87
Social Babble	506	3	8	17	63	23	59	47	130	96	60
All Speech	442	0	0	0	0	1	0	73	245	53	70
Social Speech	418	0	0	0	0	1	0	54	241	53	69
All Other Emoting	3433	350	306	617	549	412	515	254	242	107	81
Social Other Emoting	2484	210	160	381	381	262	470	241	229	85	65

Table 4-3. Sample sizes, per situational and monthly subsets of the dataset. Subsets highlighted in yellow were retained for analysis.

For each sample, feature sets were computed from:

(1) The child vocalization

(2) Adult speech ***before*** the vocalization

(3) Adult speech ***after*** "

(4) Adult speech ***surrounding*** "

(5) All of the above combined

Modeling Alternatives

- Multivariate Linear Regression
  - Supports only univariate output
  - Susceptible to overfitting

- Principal Components Regression (PCR)
  - Internal factor analysis (PCA) of input variables
    - Reduces dimensionality
    - Guards against overfitting

- Partial Least Squares Regression
  - Internal factor analysis <u>includes output variables</u>
  - Superior to PCR
    - Classification accuracy
    - Generalizability
    - Robustness to overfitting, sample size effects

PLS Regression Procedure

- Matlab: **plsregress(X, Y, nComp, ‘cv’, 10)**

  X = matrix of input variables

  Y = matrix of output variables

  nComp = # PLS components to retain

  - 10-fold cross validation

- Procedure

  1. Derive optimal nComp
  2. Apply optimal nComp to get final model
  3. Evaluate Performance

Deriving Optimal nComp

1. Run **plsregress** with nComp ≥ 50
2. Plot Mean Squared Error (MSE) vs. # PLS Components
3. Minimum MSE gives optimal nComp

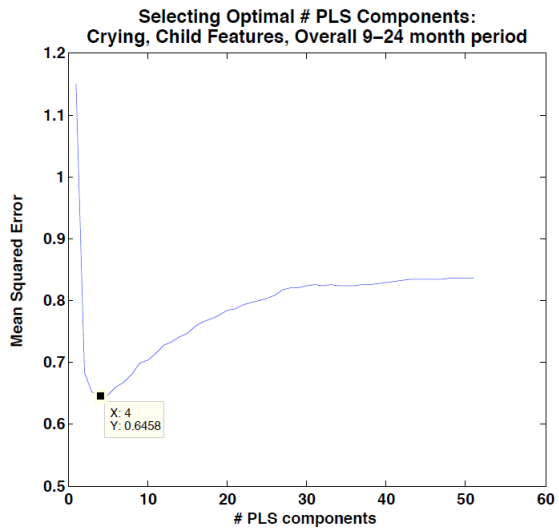


Figure 4-1. Deriving the optimal number of PLS components for a model. This particular example is for PLS regression applied to the Crying subset of child vocalization features, covering the overall 9-24 month period. A clear minimum MSE value occurs at 4 PLS components, beyond which MSE starts to rise. We therefore build the PLS model for this subset of data using 4 PLS components.

Measuring Model Performance

- R-squared
  - How well the regression line approximates the data

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \left(\sum_{j=1}^{p} x_{ij} B_j\right)\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

- Adjusted R-squared
  - Normalizing for sample size and # Components

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-q-1}$$

n	# samples
p	# input variables
B_j	Regression Coefficient for feature j
Y_i	Vector of multidimensional output values for sample i
x_{ij}	Input value for feature j in sample I
$\overline{Y}$	Sample mean of output data
q	# Components

Interpreting R^2_{adj}

Effect size	Adjusted R-squared Range
Small	$0.02 \leq R^2_{adj} < 0.13$
Medium	$0.13 \leq R^2_{adj} < 0.26$
Large	$0.26 \leq R^2_{adj}$

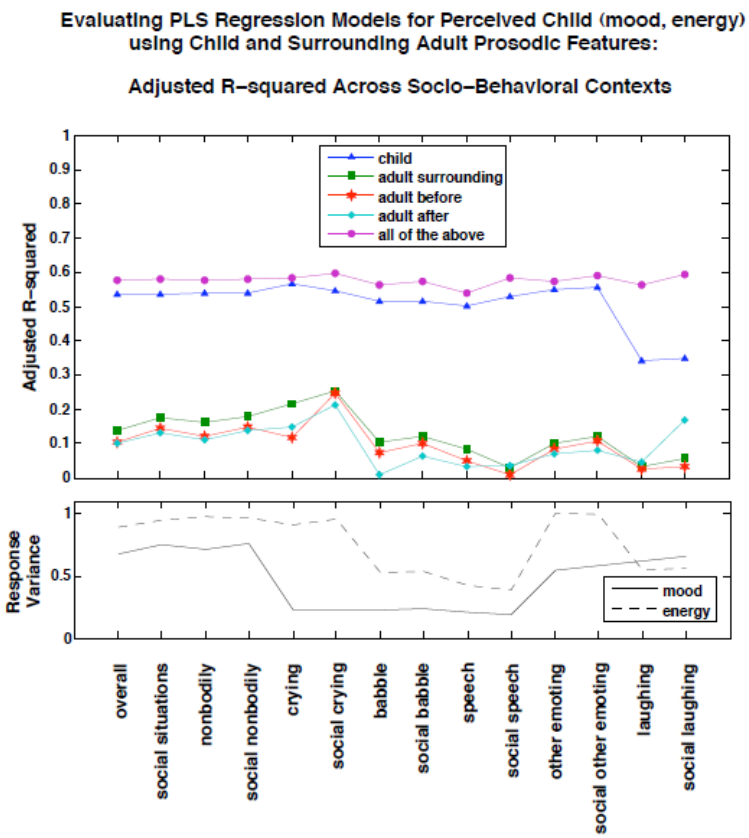
Cohen, J. (1992) A power primer. *Psychological Bulletin 112, 155-159.*

Harlow, L. L. (2005) *The Essence of Multivariate Thinking.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

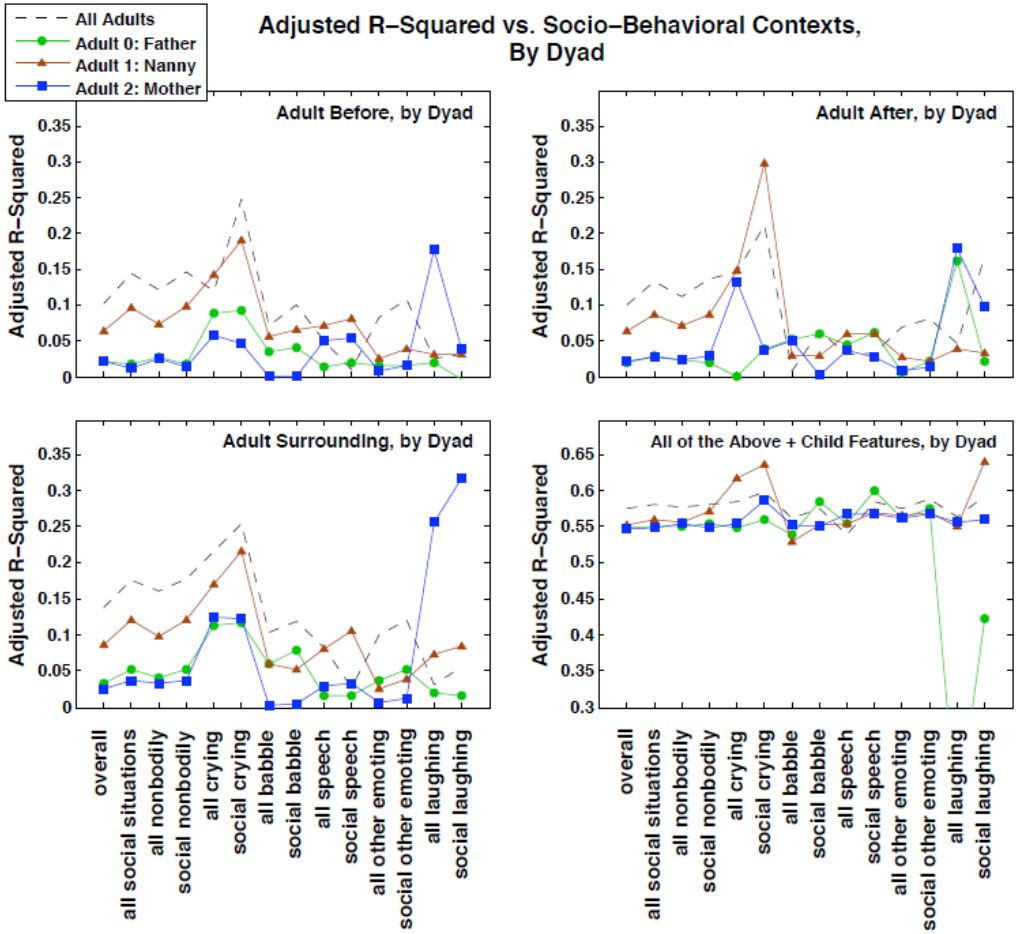
Results

Results by Socio-Behavioral Context

- ***Child*** & ***Combined***: 0.54 to 0.6
  - High performing models
    - Very large effect size >> 0.26
  - Consistent across socio-behavioral contexts
    - Social situations bring out slight increase
    - Child laughing outlier: a mystery
- Adult speech
  - Matters…
    - Medium effect size (≥ 0.13) in many cases
  - …but not too much!
    - Small effect sizes elsewhere
    - ***Combined*** improves only slightly over ***child***.

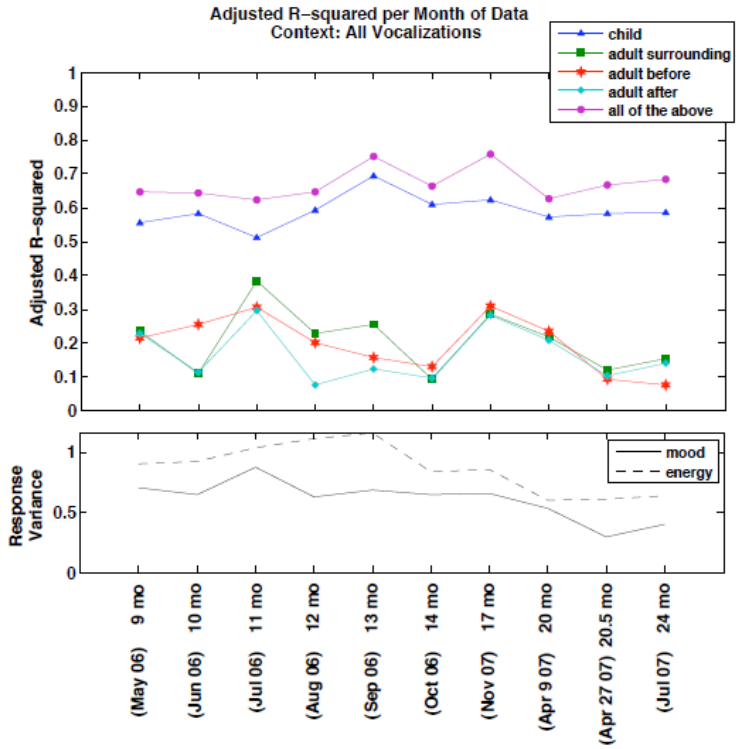


Dyadic Comparisons

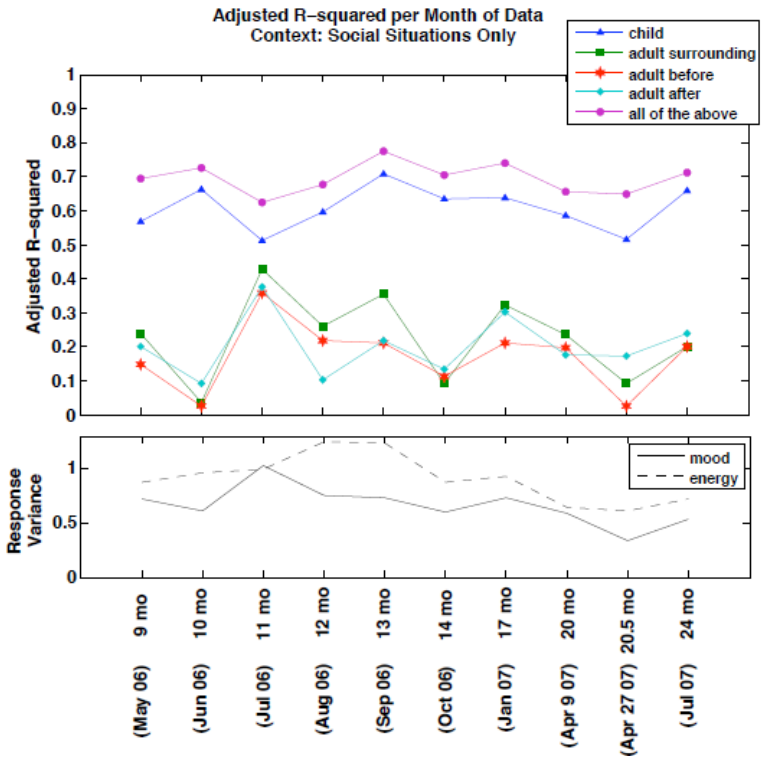


Results by Month

All Vocalizations



Social Situations Only



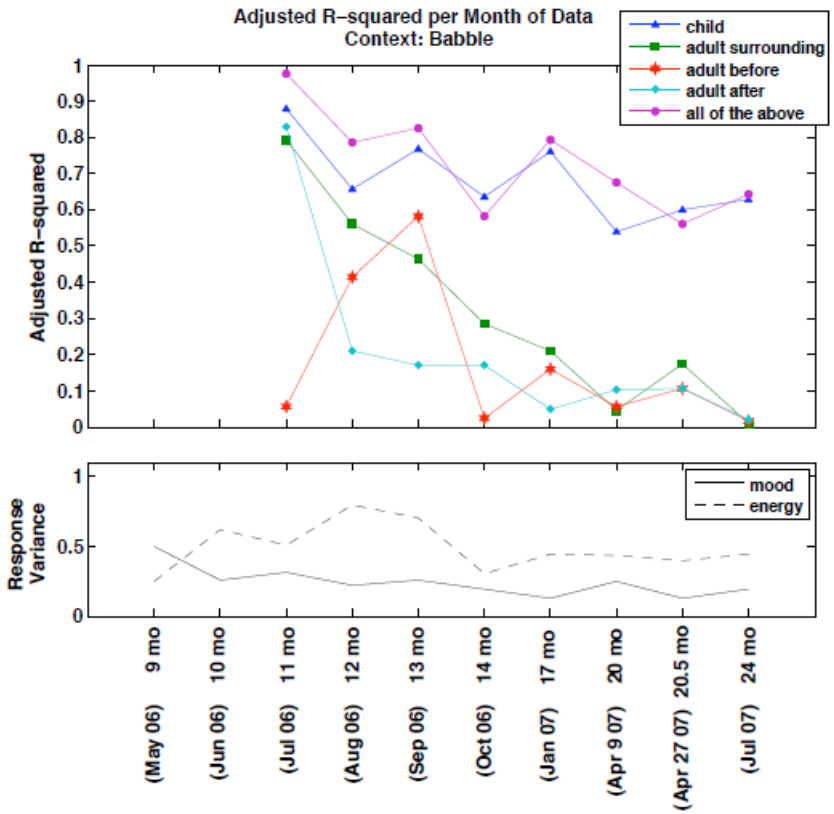
Better to Build Monthly Models

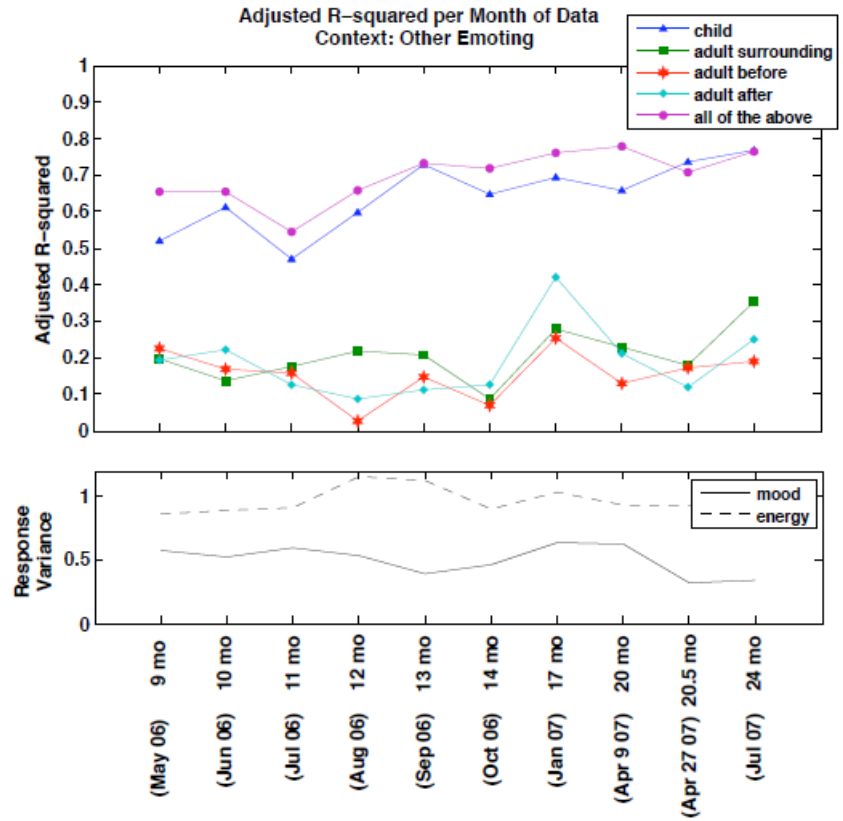
On average, monthly models outperform time-aggregate models

Context	Combined				Child				Adult Surrounding			
	Overall (Fig 5-1)	Longitudinal			Overall (Fig 5-1)	Longitudinal			Overall (Fig 5-1)	Longitudinal		
		mean	min	max		mean	Min	max		mean	min	max
All Vocalizations	0.58	0.67	0.62	0.76	0.54	0.59	0.51	0.69	0.14	0.21	0.09	0.38
Social Only	0.58	0.70	0.62	0.78	0.54	0.61	0.51	0.71	0.18	0.23	0.04	0.43
All Nonbodily	0.58	0.67	0.63	0.75	0.54	0.60	0.52	0.70	0.16	0.22	0.09	0.38
Social Nonbodily	0.58	0.70	0.63	0.78	0.54	0.62	0.52	0.71	0.18	0.24	0.04	0.48
All Crying	0.59	0.78	0.31	1.00	0.57	0.76	0.61	0.97	0.22	0.40	0.09	1.00
Social Crying	0.60	0.86	0.71	1.00	0.55	0.74	0.61	0.98	0.26	0.51	0.13	0.81
All Babble	0.56	0.73	0.56	0.98	0.52	0.68	0.54	0.88	0.10	0.32	0.01	0.79
Social Babble	0.57	0.82	0.60	0.99	0.52	0.66	0.53	0.84	0.12	0.26	0.00	0.60
All Speech	0.54	0.71	0.58	0.87	0.50	0.63	0.49	0.78	0.08	0.18	0.06	0.39
Social Speech	0.59	0.73	0.57	0.95	0.53	0.64	0.48	0.80	0.03	0.18	0.04	0.39
All Other Emoting	0.57	0.70	0.54	0.78	0.55	0.64	0.47	0.77	0.10	0.21	0.09	0.35
Soc Other Emoting	0.59	0.75	0.55	0.92	0.56	0.66	0.43	0.78	0.12	0.23	0.03	0.38

Table 5-2. Comparing Overall Performance of Month-by-Month models with Time-Aggregate models. For each Socio-Behavioral context, adjusted R-squared is significantly higher on average when using month-specific models than when using a time-aggregate model built using data from the overall 9-24 month period.

Interesting Progressions





Future considerations

- Do we really need meticulously annotated vocalizations?
  - Repeat analysis with noisy auto-detected "speech" segments
- Do the results/trends generalize?
  - Children can be more or less inhibited
  - Repeat analysis with forthcoming new Speechome datasets

Figure 6-1. Speechome Recorder

Recommended Reading

- Geladi & Kowalski (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185: 1-17
- Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.). Multivariate Analysis. (pp.391-420) New York: Academic Press.
- De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18: 251–263
- http://mathworld.wolfram.com/LeastSquaresFitting.html
- Matlab tutorial comparing PCR and PLS:
  - http://www.mathworks.com/products/statistics/demos.html?file=/products/demos/shipping/stats/plspcrdemo.html

Thank you!