



Predicting Prosodic Boundaries for Children's Texts

Mansi Dhamne¹, Sneha Raman², Preeti Rao²

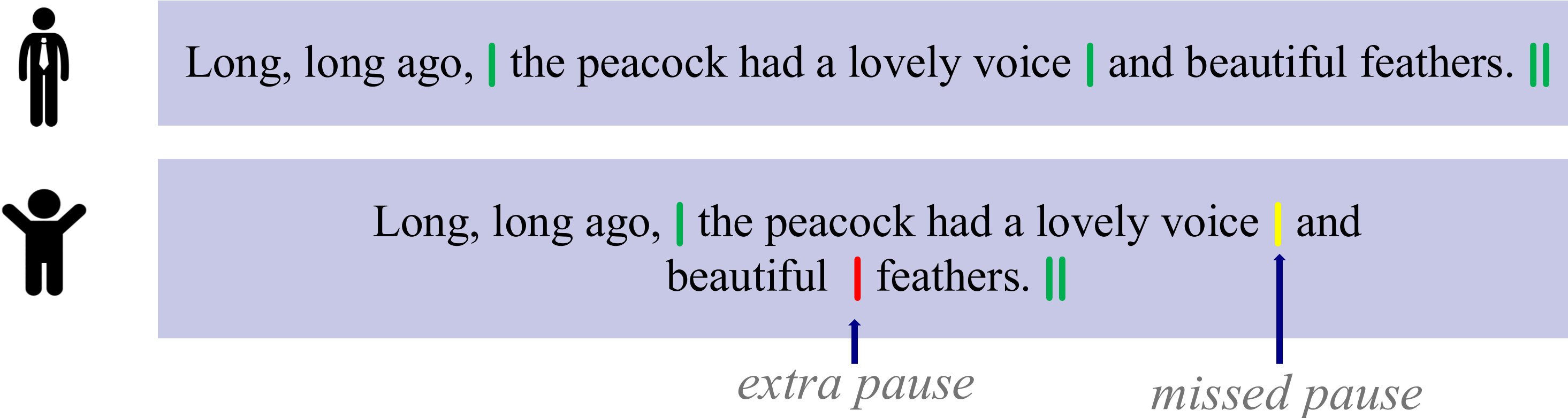
¹Department of Computer Engineering, Sardar Patel Institute of Technology

²Department of Electrical Engineering, Indian Institute of Technology Bombay



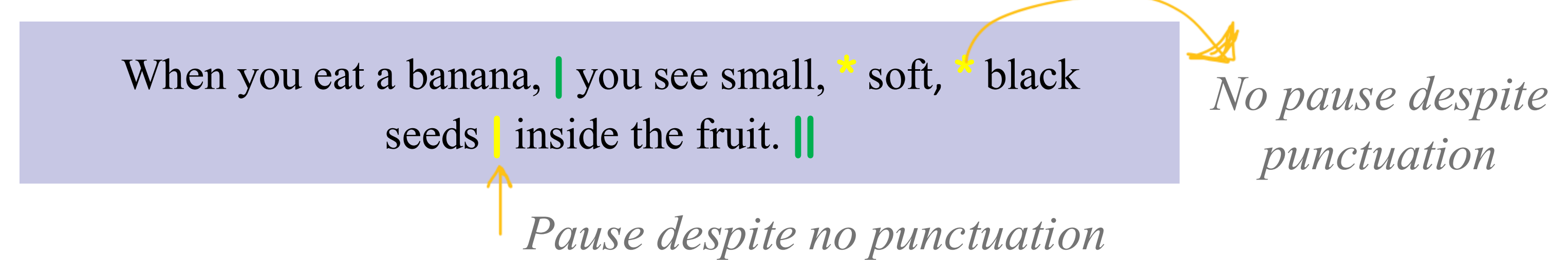
I. MOTIVATION

- Reading fluency = **decoding words** + **natural prosodic phrasing**
- Children (8–13 yrs) insert **unnecessary pauses**
→ limited breath control + under-developed reading fluency



II. CHALLENGE

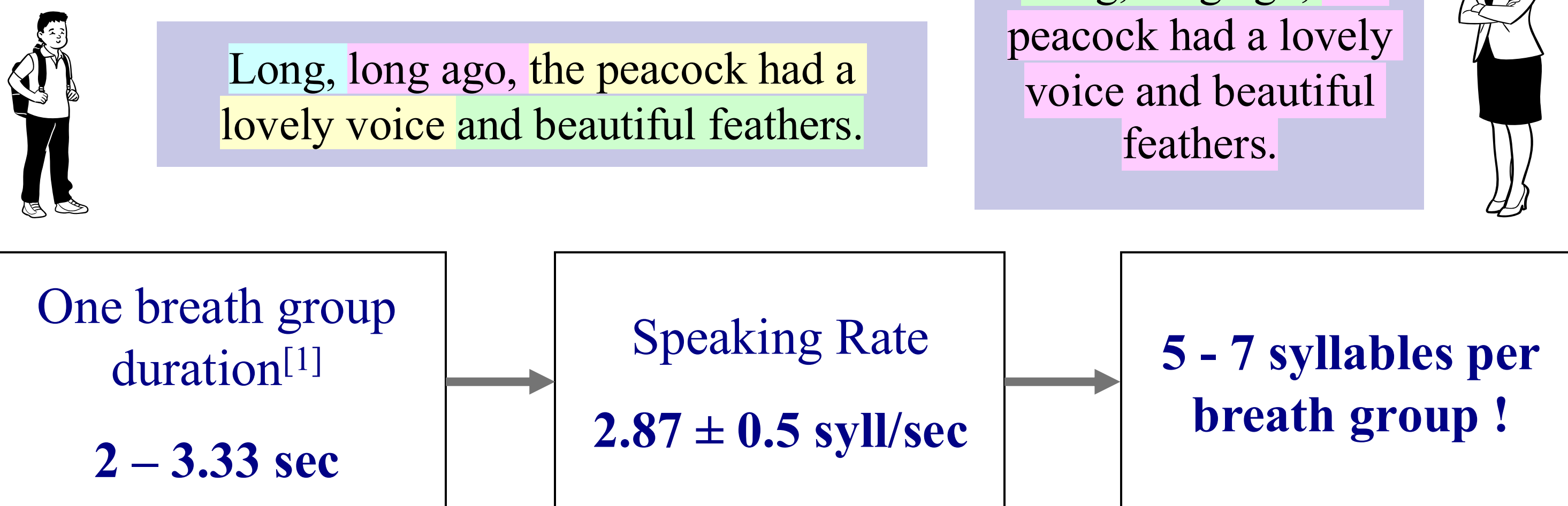
- Prosodic phrasing often **diverges from syntax** (semantic emphasis, discourse structure)



- Prosodic phrasing is **hierarchical**, not binary → multiple break levels must be modelled.
- Most previous work relies on **punctuation, acoustics or adult speech** → unsuitable for **children's speech** with cognitive & breath constraints

III. CONTRIBUTIONS

A. Breath Duration Feature



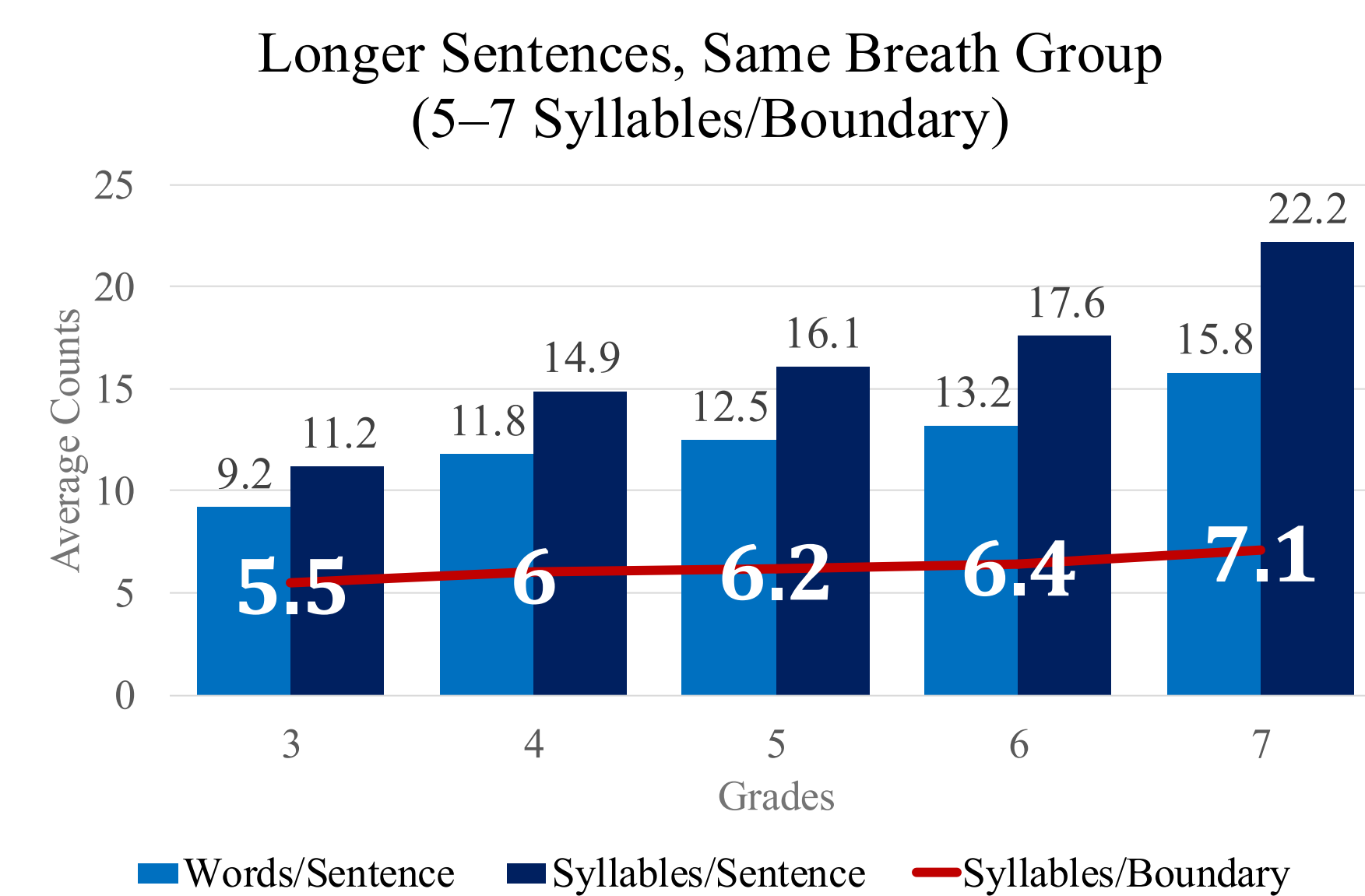
We propose cognitively motivated features:

- **Number of syllables** in the words
- **Syllables since the last major boundary**
- **Number of syllables remaining in that breath group duration** (taken as 7 syllables)

B. Human Annotated Dataset for Children's Stories

- 54 stories × 21 annotators (7 per batch)
- Annotators marked where they would naturally pause when reading aloud to children.
- Average Fleiss' Kappa = **0.68** → **Substantial agreement**

Annotated Data Insights



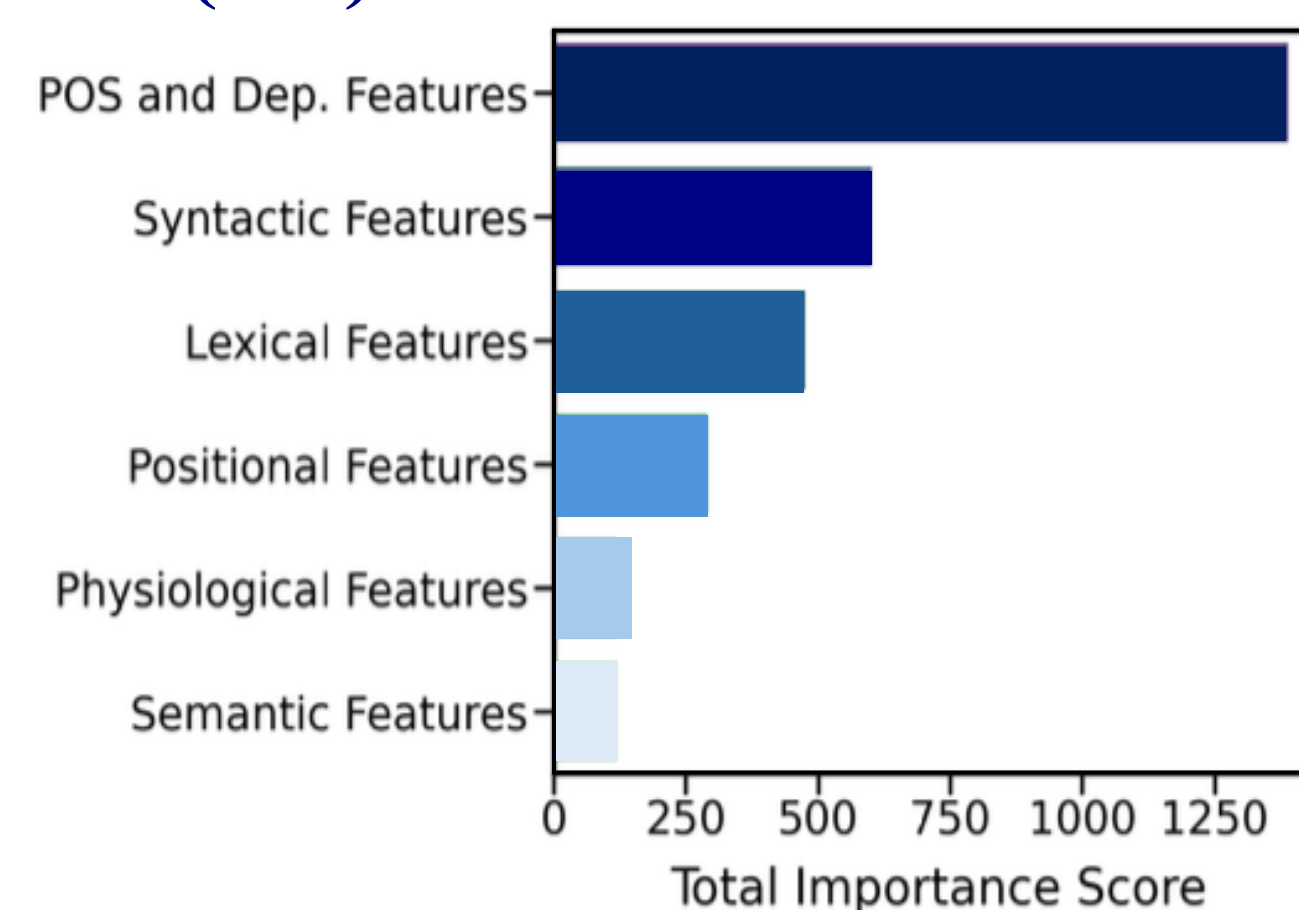
- Sentences get **longer with grade**
- Consistent with **child breath group limits**
- Only **57% agreement** for non-terminal (comma, -) punctuations.
- **Clause boundaries** (esp. “and”, “but”) align strongly with pauses.

IV. EXPERIMENTAL TASKS

Task 1: Boundary Strength Prediction

Model: LGBM Regressor; GT: **Count of Votes (0-7)**

Feature Set	R ²	RMSE	MAE
Punctuation	0.67	1.48	1.05
+ Lexical	0.70	1.44	0.95
+ Semantic	0.76	1.27	0.85
+ Syntactic and POS	0.79	1.16	0.75
+ Positional	0.80	1.14	0.73
+ Physiological	0.87	0.94	0.57



Task 2: Prosodic Boundary Detection

Model: LGBM Classifier; GT: **1 if count of votes >= 5 else 0**

Performance Metrics	PBD
Precision	0.92
Recall	0.85
F1-score	0.89
Accuracy	0.96

Task 3: Forbidden Pause Detection

Model: LGBM Classifier; GT: **1 if count of votes = 0 else 0**

Performance Metrics	FPD
Precision	0.83
Recall	0.91
F1-score	0.87
Accuracy	0.84

- ⚠ **Overlooks expressive or intonational pauses** that lie beyond strict syntactic rules or punctuation.
- ⚠ **Struggles with context prominence or subtle shifts in information structure.**

V. EXTRACTED FEATURES

- **Lexical:** Word length, word frequency, function vs content word
- **Semantic:** Word embeddings-based clustering
- **Syntactic/POS-based:** POS, phrase boundaries, dependencies
- **Positional:** Sentence/paragraph position, distance to punctuation
- **Physiological:** Breath group duration using syllables

VI. FUTURE WORK

- Child-based TTS and ORF assessment applications
- Expand to languages like Hindi and Marathi

REFERENCES AND ACKNOWLEDGEMENTS

- [1] Susannah Fleming, et. al. 2011. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies.
- [2] Rosenberg, Andrew. Automatic detection and classification of prosodic events. Columbia University, 2009.
- We thank IRCC, IIT Bombay, for supporting this research through the first author's internship program.