Relating Intonational Pragmatics to the Pitch Realizations of Highly Frequent Words in English Speech to Infants

Carolyn Quam
Jiahong Yuan
Daniel Swingley

CogSci conference, 24 July 2008

Finding the relevant variation

- Speech varies in ways relevant and irrelevant to word learning
  - "Ball" vs. "doll": important to distinguish
  - "Ball" spoken by Mom vs. Dad is not
- How does the child figure out which features distinguish words?

Pitch variation

- Pitch contrasts words in many languages
- In English, it's relevant at other levels of structure
  - Marking yes/no questions
  - Conveying the speaker's emotions
- How does the input tell children that pitch doesn't contrast words in English?
  - Consistent pitch: suggests pitch is part of the word
  - Variable pitch: suggests it's not

Pitch realizations of English words

- English is not a tone language
  - We expect variability in pitch for individual words
- **But** the simplicity of infant-directed speech might lead to consistency in a word's pitch
  - Exaggerated intonation
  - Short, simple phrases
  - Small inventory of emotional & pragmatic meanings

Motivation

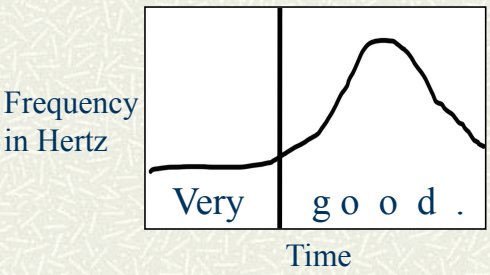
- If English words have consistent pitch, this would pose a learning puzzle
- How does the speech infants hear tell them that pitch does not differentiate words in English?
- We look at the pitch patterns of highly frequent words
  - How consistent is their pitch across tokens?
  - What influences their pitch patterns?

The Brent corpus (Brent & Siskind, 2001)

- 16 American mothers' speech to their 9- to 15-month-olds
- Roughly 200 hours of naturalistic interaction (about 400,000 words)
- Transcription divides the corpus into utterances

Getting the pitch contours of **words**

- Time stamps in Brent transcription let us locate each utterance in the sound files
- For each utterance, we extracted fundamental frequency in Hertz (perceived as pitch)
  - Excluded pitch-sample outliers
  - Conversion from Hertz to the Mel scale
  - Z-score normalization for each speaker



- To find the words, we used **HTK forced alignment**
  - Automated method of splitting the sound file up into words

Analyzing a subset of tokens

- Eight highly-frequent words
  - Good, right, no, okay, up, down, ball, & book
- Excluded noisy, whispered, or sung tokens
- Tokens in *final position*
  - Where word's pitch is most likely to be realized fully
  - Infants recognize words better in final position
- Tokens in *statements*
  - Avoids effects of sentence-level intonation

Typical contexts for each word

Good: "...very good" (106 tokens); "...so good" (46);
"...that's good" (36); "...mmmm good" (29); "...it's good" (27).

Right: "...that's right" (464); "you're right" (15).

Both have approving function

Typical contexts for each word

Good: "...very good" (106 tokens); "...so good" (46);
"...that's good" (36); "...mmmm good" (29); "...it's good" (27).

Right: "...that's right" (464); "you're right" (15).

No: "...no no" (607); "...oh no" (133).

Okay: "...it's okay" (147); "...you're okay" (41); "...that's okay" (32).

No: mostly prohibitive

Okay: mostly comforting

Typical contexts for each word

Good: "...very good" (106 tokens); "...so good" (46); "...that's good" (36); "...mmmm good" (29); "...it's good" (27).

Right: "...that's right" (464); "you're right" (15).

No: "...no no" (607); "...oh no" (133).

Okay: "...it's okay" (147); "...you're okay" (41); "...that's okay" (32).

Up: "...it up" (60); "...you up" (54); "...stand up" (15); "...clean(ed) up" (23).

Down: "...fall/fell down" (57); "...sit down" (30); "...upside down" (20); "...get down" (17); "...up and down" (11)

Both prepositions, but opposite meanings

Typical contexts for each word

Good: "...very good" (106 tokens); "...so good" (46); "...that's good" (36); "...mmmm good" (29); "...it's good" (27).

Right: "...that's right" (464); "you're right" (15).

No: "...no no" (607); "...oh no" (133).

Okay: "...it's okay" (147); "...you're okay" (41); "...that's okay" (32).

Up: "...it up" (60); "...you up" (54); "...stand up" (15); "...clean(ed) up" (23).

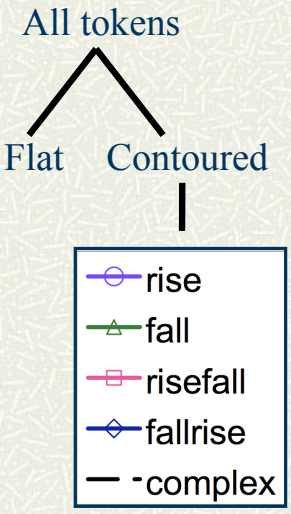
Down: "...fall/fell down" (57); "...sit down" (30); "...upside down" (20); "...get down" (17); "...up and down" (11)

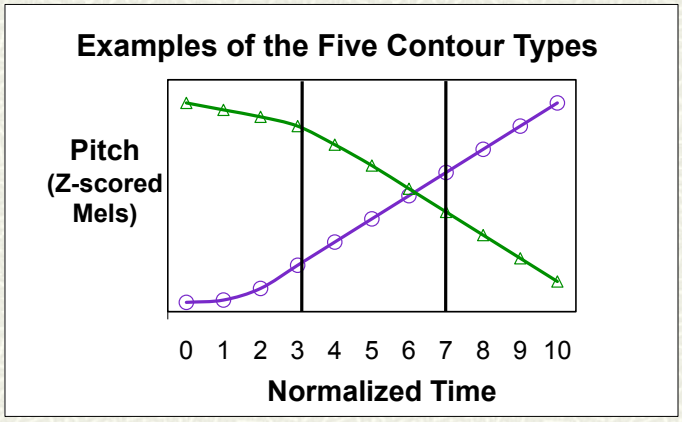
Ball: "...the ball" (98); "...your ball" (35).

Book: "...this book" (32); "...the book" (28); "...a book" (25); "...your book" (23).

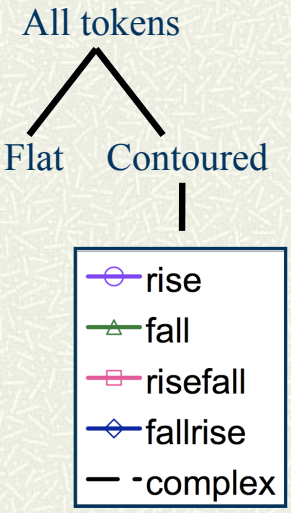
Both concrete nouns

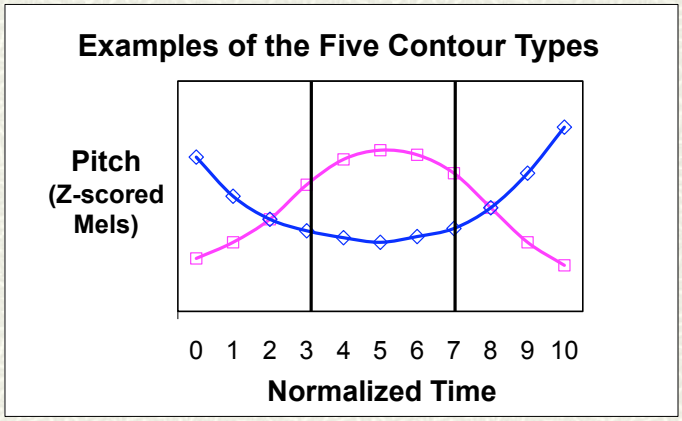
Assigning each token a contour type



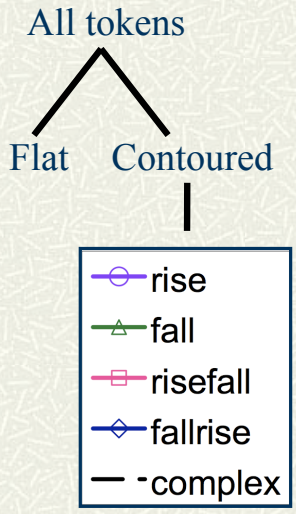


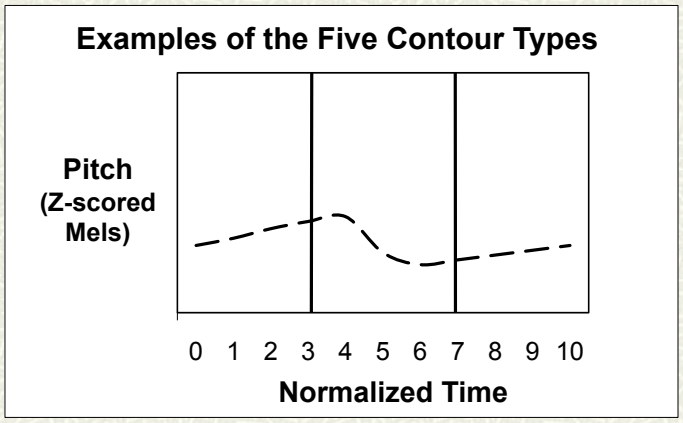
Assigning each token a contour type



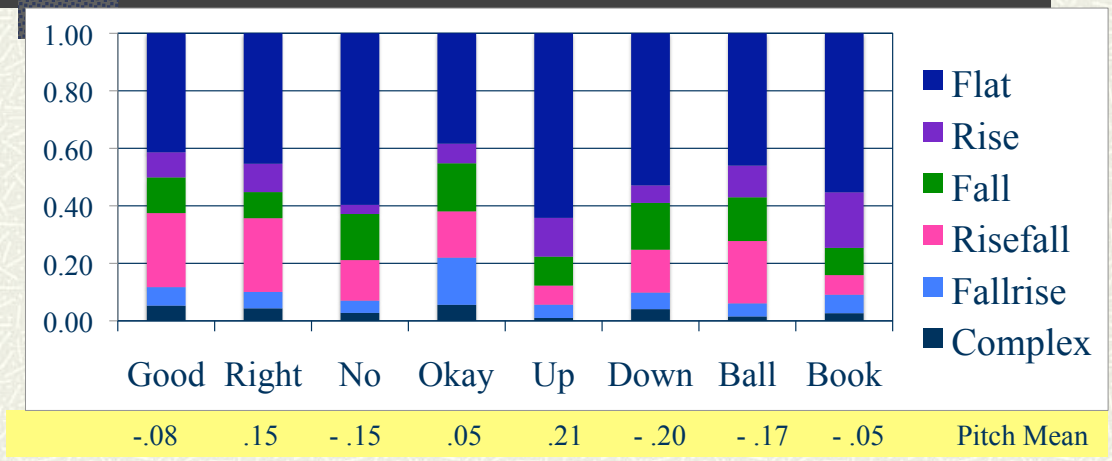


Assigning each token a contour type



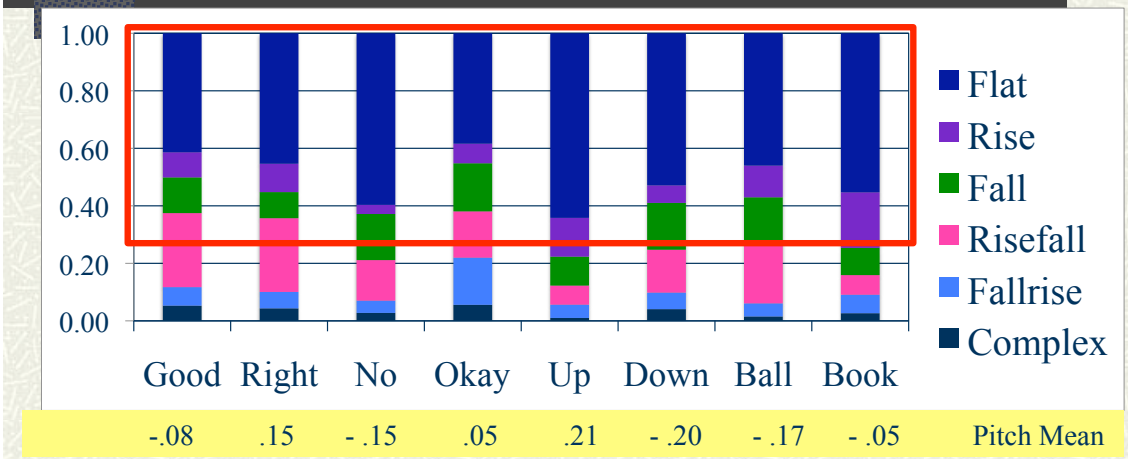


Final position in statements



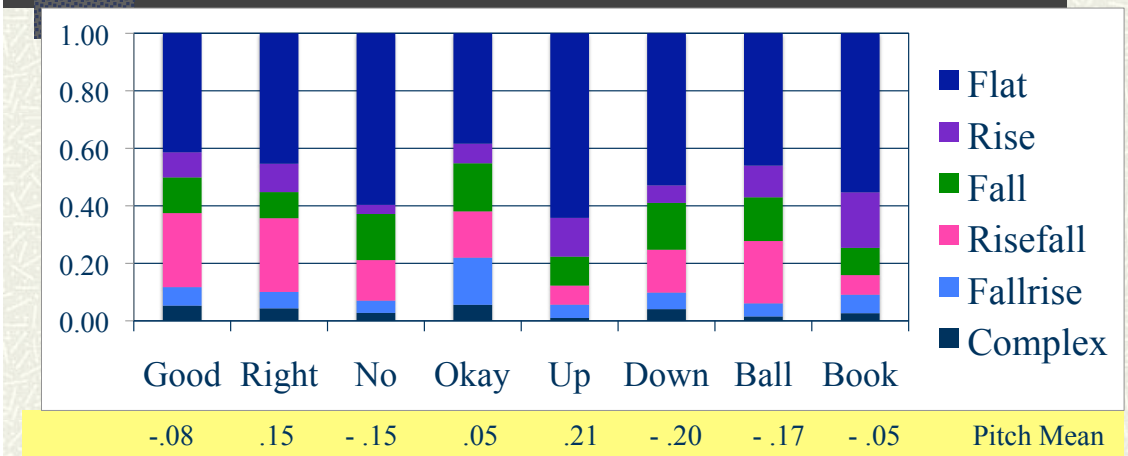
Contour-type distributions and pitch means

Final position in statements



- Lots of flat tokens
  - Surprising, since exaggerated pitch movement is the hallmark of infant-directed prosody

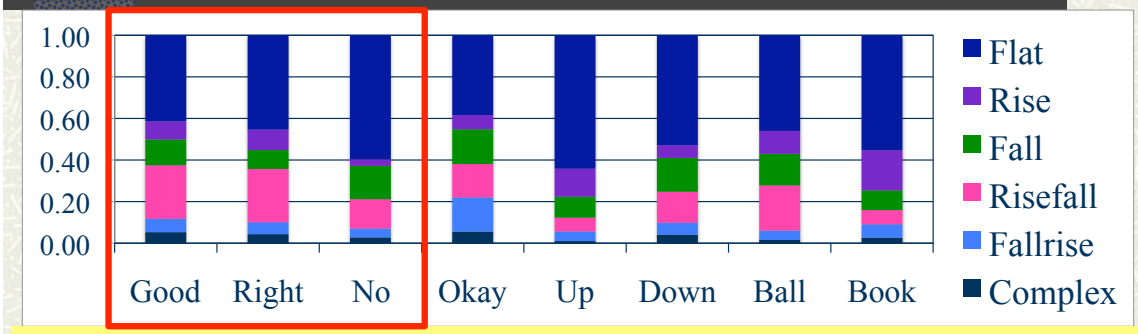
Final position in statements



Contour Similarity $_{Good/Right}$ = sum(differences)

Difference $_{Fall}$ = Abs(Good $_{Fall\%}$ - Right $_{Fall\%}$)

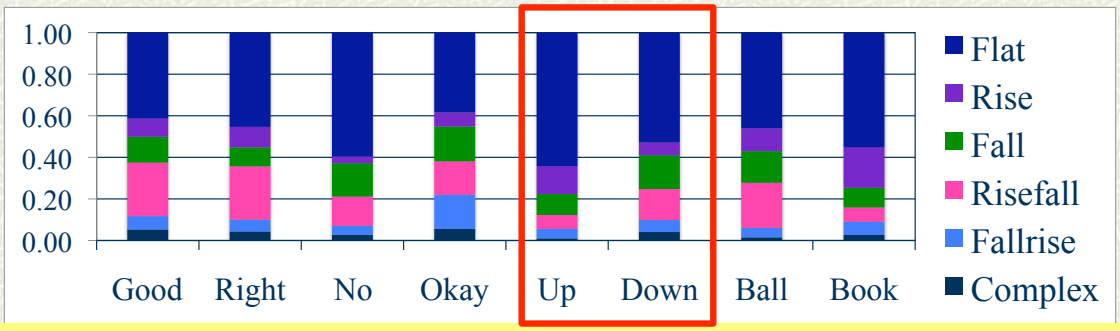
Good & right versus *no*



- .08 .15 - .15 Pitch Mean

- *Good & right* have the highest contour similarity
- *Good & no* and *right & no* are much less similar
  (23rd of 28) (21st of 28)
- *Right* and *no* also differ in pitch mean ($p < .05$)

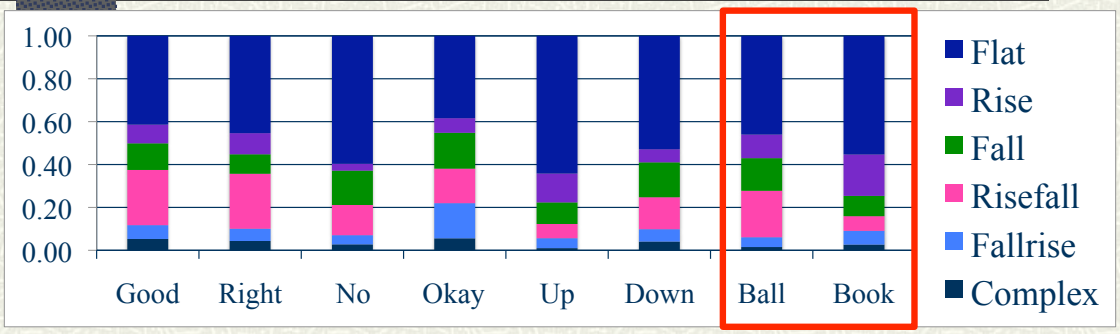
Up vs. *down*



.21 - .20 Pitch Mean

- *Up* and *down* are below average in contour similarity (16th of 28)
  - *Up* has more rises ($p = .13$), *down* has more falls ($p = .14$)
- *Up* has a higher pitch mean ($p < .001$)

Ball vs. *book*

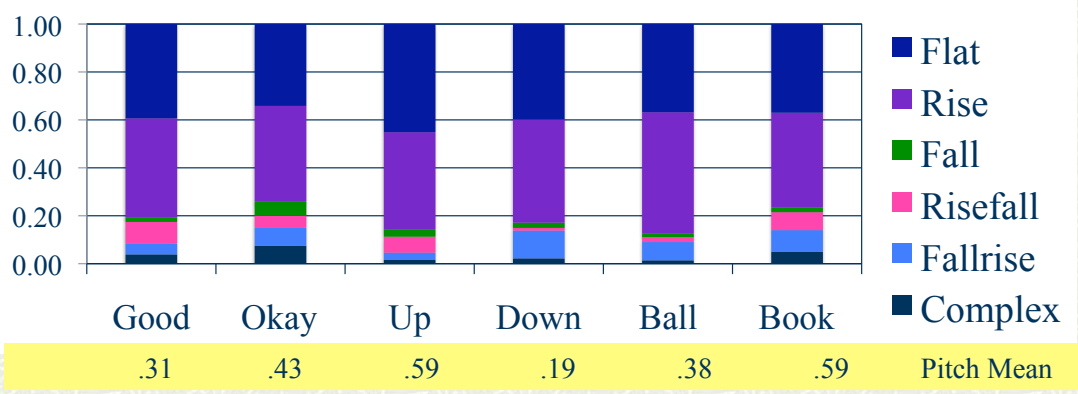


- Both words are concrete nouns, but they're below average in contour similarity (19th of 28)
- No evidence that pitch indicates category *noun*

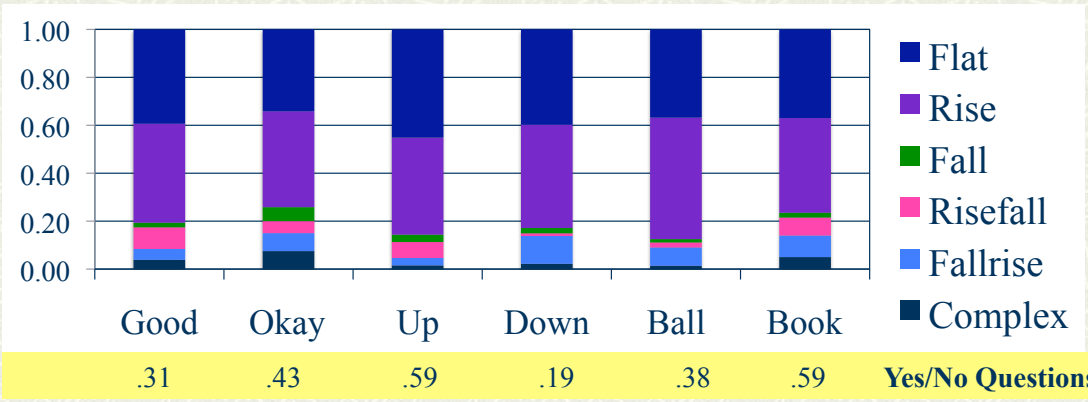
Discussion

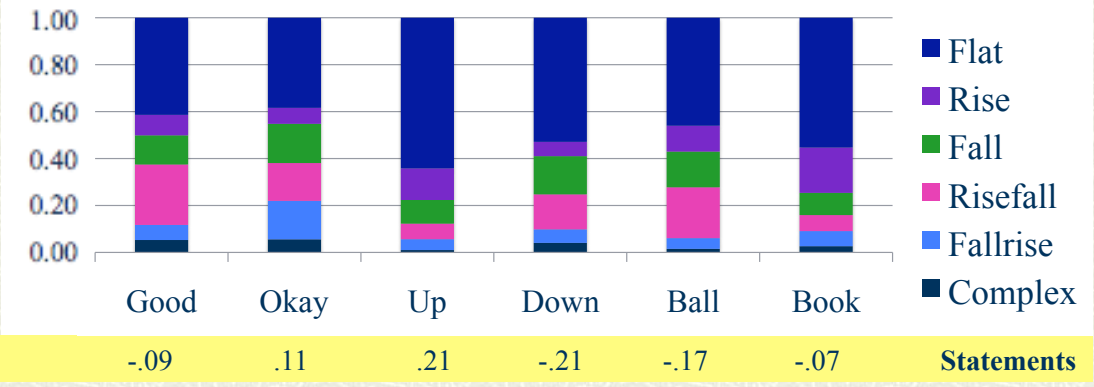
- Though words are variable, some words differ in their pitch patterns
- But how large are these differences?
- We can calibrate the differences between words against a function we know pitch plays in English
  - Marking yes/no questions

Yes/no questions



- Every word has more rises than it does in statements ($p < .001$ for each test)
- And higher pitch means ($p < .001$ for each test)





Discussion

- Children must attach the salient pitch movements they hear to some level of structure
  - Mostly, they don't seem to indicate particular words
  - They perform other functions, like indicating pragmatic function and marking yes/no questions

Conclusions

- But what about the differences between words?
  - Large within-word variability may convey that English is not a tone language
  - And the differences between words seem to mostly reflect pragmatic functions
    - E.g., *good* and *right* are used approvingly, and have similar pitch patterns
    - *No* is used in prohibitions, and has different pitch patterns

The importance of corpus analyses

- Knowledge of phonological development comes mostly from experimental work
  - Demonstrates children's knowledge of native language sounds—but how do they learn them?
- Corpus analyses characterize the complex input to children
  - Important if we want an accurate view of the language-learning problem

Acknowledgements

We thank Kyle Gorman, Mark Liberman, Stephen Isard, Chandan Narayan, Delphine Dahan, Frank Newman, Marisa Macias, and Brian Decker for helpful comments and technical support. We also thank members of the Phonetics lab and IRCS for their valuable suggestions. Funding was provided by NSF Graduate Research Fellowship and NSF IGERT Trainee Fellowship grants to C.Q., and NIH grant R01-HD049681 to D.S.

Contact

cquam@psych.upenn.edu

(http://wordle.net/)