



*It's the Difference That Matters:*  
**An Argument for  
Contextually-Grounded Acoustic  
Intonational Phonology**

Sasha Calhoun  
Centre for Speech Technology Research  
University of Edinburgh, UK  
[Sasha.Calhoun@ed.ac.uk](mailto:Sasha.Calhoun@ed.ac.uk)

*Linguistics Society of America Annual Meeting  
Oakland, CA, 6 January 2005*





# Overview

- what makes a good intonational phonology?
- issues with ToBI: description, explanation, verification
- pilot study: using linear regression modelling to find acoustic cues to topic status
- intensity, duration, phrase level and relative  $f_0$  cues all significant
- suggest phonological investigations should be corpus-based, categories are bundles of weighted acoustic features affected by context





# Qualities of a good phonology

= the structure of supra-segmental speech sounds

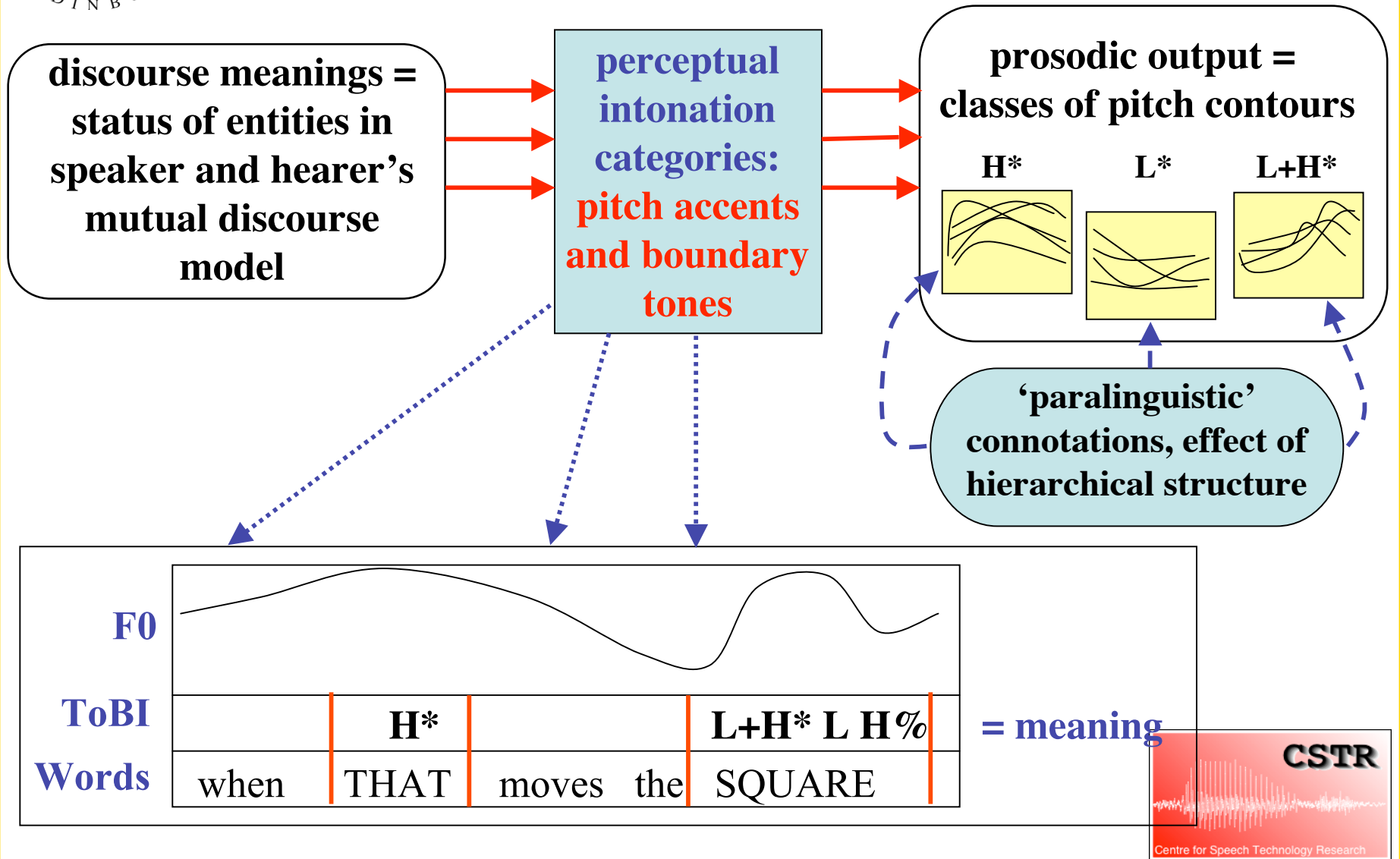
- **describe** parts of speech signal relevant to the conveyance of intonation categories
- **explain** how these intonation categories convey meaning
- be **verifiable**
- give **coverage** of differences between languages and varieties of one language





# Intonational Phonology & ToBI

(Silverman et al 92, Pierrehumbert & Hirschberg 90)





# Issues

## ■ description:

- bias on **local features** of pitch contour (c.f. importance of *relative* height: Gussenhoven & Rietveld 88, Terken 91, Ladd 96;  $f_0$  on unstressed syllables: Xu et al 04) →
- bias on  **$f_0$  turning points** (c.f. other factors which affect alignment: van Santen & Möbius 00, segmental effects; Scherer et al 04, emotions) →
- no explicit modelling of effect of other **layers of structure**

- **linear regression modelling**
- **bundles of acoustic features**
- **include previous context as features**



# Issues

## ■ explanation:

- very **uneven distribution** of pitch accents (Taylor 00: 80% H\*)
- little evidence of emerging consensus on **pitch accent meanings** (e.g. status of L+H\*, H\* P&H 90, Steedman 00, Lambrecht & Michaelis 98 all differ, Hedberg & Sosa 01 corpus - mixed)

→

■ **start with meanings**

## ■ verification:

- **low inter-annotator agreement** on pitch accent types (Silverman et al 92: 61%)
- difficult to find criteria to confirm or reject existence of **perceptual categories**

→

■ **corpus-based research**





# Topic Status Experiment with SPOT Corpus

- investigate intonational marking of topics in SPOT corpus
- tested given, new and contrastive categories
- SPOT = dialogues collected as part of a game task by Schafer (Hawaii), Speer (Ohio), Warren (Victoria, NZ) and colleagues
- used 52 utterances involving 16 pairs of male speakers of American English
- these utterances ToBI transcribed as part of the original project





# Discourse Contexts

- tested the realisation of the word *square* in different discourse contexts in a game task where people had to, among other things, move squares with cylinders.

Q: Which cylinder do you want to change the position of the square?

A: The red one. When that moves the *square*, it should land in a good spot.

*given topic*

Q: Which cylinder do you want to change the position of this time?

A: The red one. When that moves the *square*, it should land in a good spot.

*new topic*

Q: ( I know which cylinder you want to change the position of the triangle, )  
but which cylinder do you want to change the position of the square?

A: The red one. When that moves the *SQUARE*, it should land in a good spot.

*contrastive topic*







# ToBI - No clear mapping

- no statistically significant relationship between ToBI pitch accents and topic status

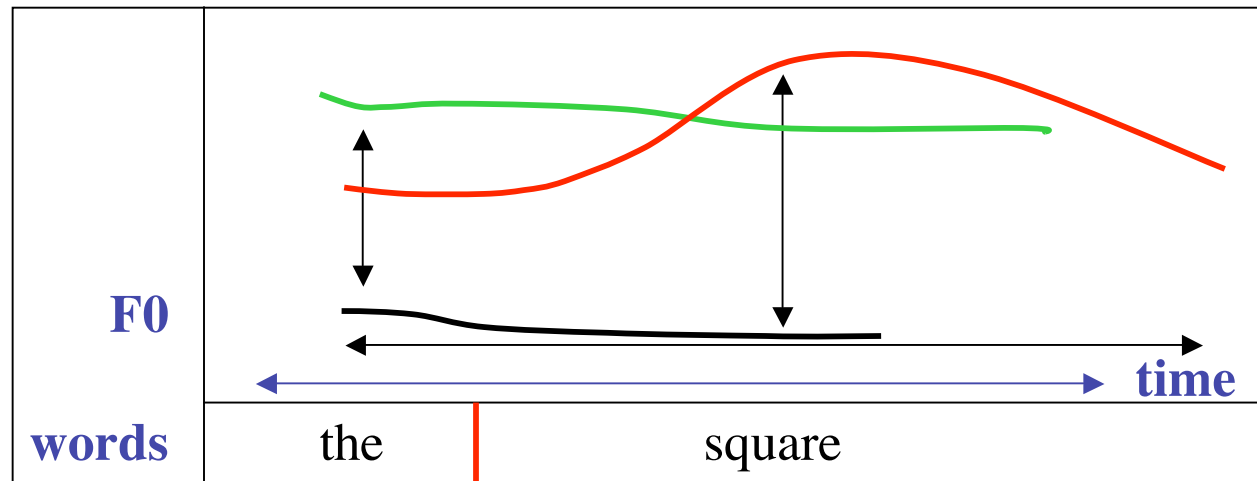
	$\emptyset$	L*	H*
Given	4 (44%)	1 (11%)	4 (44%)
New	6 (19%)	6 (19%)	19 (61%)
Contrastive	4 (33%)	1 (18%)	7 (58%)



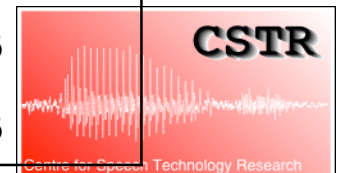
# 'Bundles' of Acoustic Cues

- using a linear regression model, the  $f_0$  mean of *the* and *square*, and the duration of *square* were all significant predictors of topic status:

**given**  
 new  
 contrastive



	<i>the</i> $f_0$ mean (semitones base 100 Hz)	<i>square</i> $f_0$ mean (semitones base 100 Hz)	<i>square</i> duration (msec)
given	-0.72	-1.11	365
new	1.71	1.23	423
contrastive	0.922	2.08	463

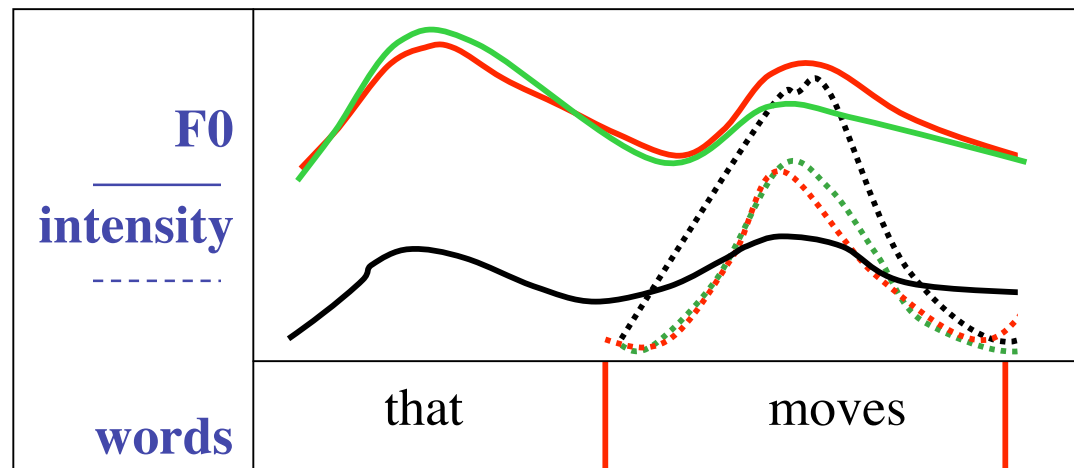




# Contextual acoustic cues

- however, we find that topic status can also be predicted using a linear regression model with acoustic features of the preceding utterance as features

**given**  
**new**  
**contrastive**



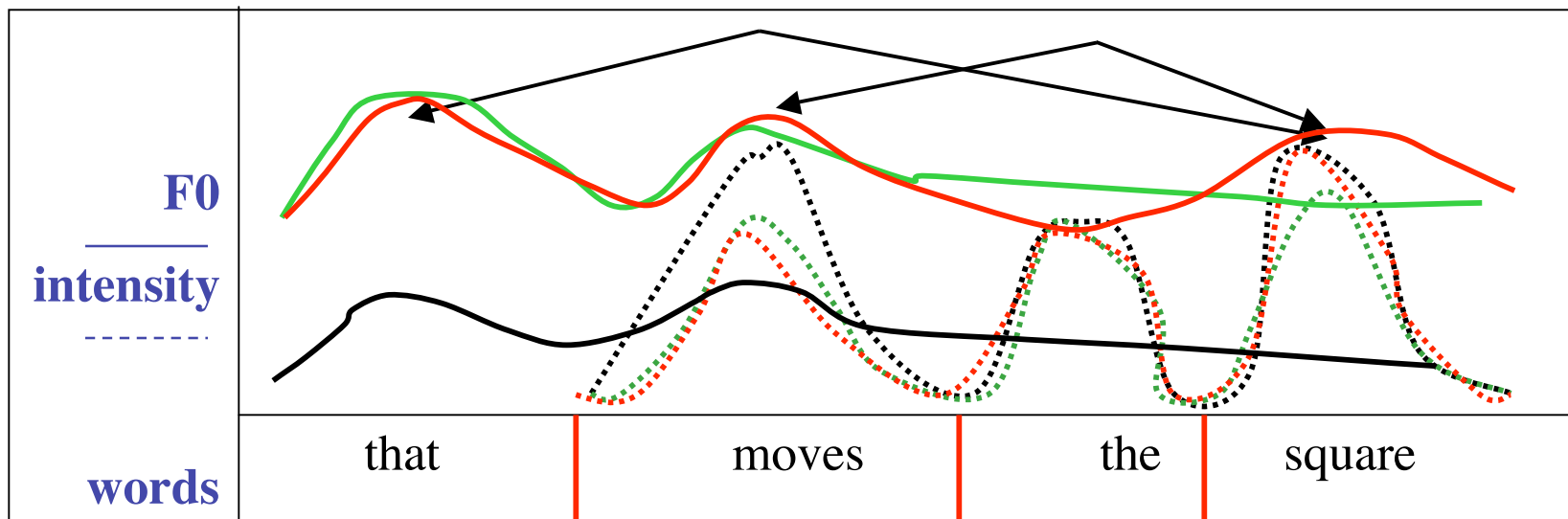
	<i>that</i> <i>f0</i> mean	<i>moves</i> <i>f0</i> maximum	<i>moves</i> mean intensity (dB)	phrase <i>f0</i> mean	phrase mean intensity (dB)
<b>given</b>	0.0	1.4	67	- 0.6	75
<b>new</b>	2.6	3.2	65	1.6	72
<b>contrastive</b>	2.5	3.6	65	2.1	74





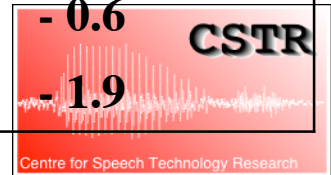
# Relative Acoustic Cues

- topic status can also be signalled by the level of different acoustic cues on *square* relative to their value in the preceding utterance.



**given**  
**new**  
**contrastive**

	<i>that - square f0 difference</i>	<i>moves - square f0 difference</i>	<i>moves - square intensity difference</i>
<b>given</b>	1.1	2.5	- 0.5
<b>new</b>	1.3	2.0	- 0.6
<b>contrastive</b>	0.4	1.5	- 1.9





# Summary - Topic Marking

- intensity and duration are significant cues to intonation categories along with  $f_0$
- ‘given’ versus ‘new’/‘contrastive’ topic status appear to be features of whole intonation phrases
- ‘new’ versus ‘contrastive’ topic status could be marked by the relative  $f_0$  height and intensity of *square* compared to *that* and *moves*



# Conclusions and Research Directions

- seeing intonational categories as bundles of weighted acoustic features allows statistical modelling of intonational phonology
- this increases **descriptive** power and makes the model **verifiable**
- the approach also **explains** how meaning is conveyed much more transparently
- statistical variation explicitly models contextual variation
- larger studies on these lines allow a more principled way to discover the apparatus of intonation, e.g. pitch accents, branching structure, phrase properties, etc.

