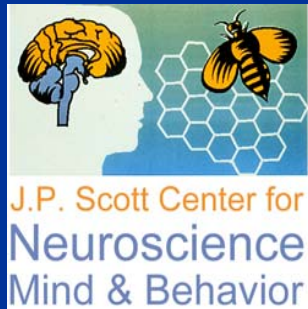


# Empirical perspectives on prosodic structure

Laura Dilley, Ph.D.

Bowling Green State University

April 11, 2008



## Thesis (N, plural theses)

1. a proposition stated or put forward for consideration, esp. one to be discussed and proved or to be maintained against objections
2. a subject for a composition or essay.
3. a dissertation on a particular subject in which one has done original research

## 4. Prosody.

- a. a part of a metrical foot that *does not bear* the ictus or stress.
- b. (less commonly) the part of a metrical foot that *bears* the ictus. Compare arsis.

-Random House Unabridged Dictionary (2006)

# Overview

- Thesis of the talk: A number of prosodic phenomena can be understood in terms of *general principles of auditory perceptual organization*
- These principles explain some cases of why listeners hear prosodic boundaries
- Implications for word segmentation

# Prosodic structure

- Prosodic structure is the organizing framework of speech (Beckman & Edwards, 1994)
- Elements of speech are grouped together into prosodic constituents
- These constituents are delimited by prosodic boundaries
  - Prosodic boundaries “set off” grouped elements, such as words, belonging to different constituents

# Organization of prosody

- Prosodic constituents are arranged according to a hierarchy (Nespor & Vogel, 1986; Beckman & Pierrehumbert, 1986)
  - mora < foot < phonological word < clitic group < phonological phrase < intonation phrase
- Boundaries which mark constituents that are higher in the hierarchy are associated with greater prosodic “strength” (Fougeron & Keating, 1997; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996)
  - A complex constellation of phonetic cues mark phrase boundaries

# Recognizing prosodic boundaries?

- Prosodic cues can be used to disambiguate syntax (e.g., Lehiste, 1972)
- Prosodic cues are useful in word segmentation
  - Speech consists of a continuous stream of acoustic material
  - Listeners posit word boundaries before stressed syllables (Cutler & Norris, 1988; Jusczyk et al., 1999) and at prosodic phrase boundaries (Gout et al., 2004; Christophe et al., 2004)
- A widespread assumption is that prosodic boundaries are perceived by identifying boundary-related phonetic cues at these locations
  - E.g., increased lengthening, glottal allophones, etc.

# Hypothesis

- *Contextual* prosodic cues can cause listeners to hear prosodic boundaries
- Method: Perception of sequences of syllables with ambiguous lexical organization: *foot note book worm*
  - Lexical boundaries correspond to prosodic boundaries at the level of the prosodic word (PWd) and higher
  - If contextual prosodic cues influence perception of *word* boundaries, then they are also influencing perception of *prosodic* boundaries

# Proximal vs. distal prosodic cues

- Previous work shows segmentation can be influenced by *proximal* prosodic cues on a syllable (cf. stress) or just before it (cf. phrase boundary)
- Q: Can segmentation be influenced by *distal* prosodic cues two or more syllables distant from the segmentation point?
  - If so, then listeners have inferred from distal context a prosodic phrase boundary at least the size of the prosodic word (PWd) or larger

# Patterns in prosodic systems

- Listeners develop expectations about stimulus structure based on patterns in preceding context (Woodrow, 1911; Jones, 1976; Cutler, 1976; Kidd, 1989; McAuley & Kidd, 1998)

..HLHLHL.. => (HL) (HL)  
(LH) (LH)

- Alternating pitch patterns and perceptual isochrony are widespread in prosodic systems

- Example: Repetition in accentual sequences

Ladd (1986):



I wanted to read it to Julia.

H H HL L HL

- Speakers tend to repeat prosodic structures in sequence (Pierrehumbert, 2000)

# Perceptual organization

- Repeating patterns in pitch and time lead to:
  - Perception of structure: grouping and meter
  - Generation of expectation of continuation

Pitch:

HLHLHL ...  $\rightarrow$  (H\* L) (H\* L) (H\* L) ...  
H (L\* H) (L\* H) (L\* ...)

Time:

○ ○ ○ ○ ○ ○ ...  $\rightarrow$  (○ ○) (○ ○) (○ ○) ...  
\* \* \*

(Woodrow, 1911; Povel & Essens, 1985; Handel, 1989)

# Perception of structure

- Listeners organize repeating sequences to have parallel grouping structure (Lerdahl & Jackendoff, 1983)



- Hypothesis: In speech, repeating patterns in pitch and time create a sense of grouping which carries over “downstream” to influence perception of word and phrase boundaries

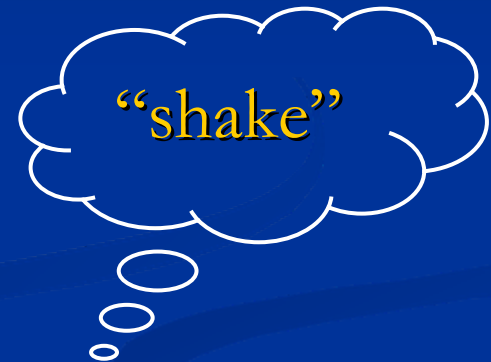
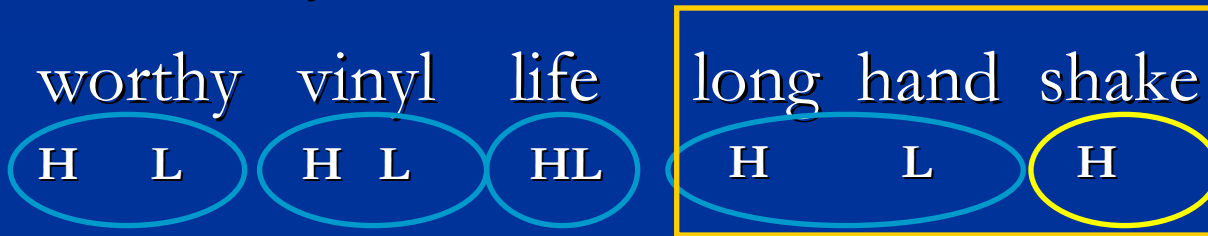
# Stimuli and Task

- 20 target sequences consisted of two disyllabic trochaic words (e.g., worthy vinyl) followed by a final four syllable string that could be organized into words in more than one way (e.g., lifelong handshake versus life longhand shake).
- 80 filler sequences with unambiguous lexical structure consisted of 6 – 10 syllables; an equal number ended with a disyllabic or monosyllabic final word.
- **Task:** Participants listened to target and filler sequences and reported the final word they heard in each sequence.

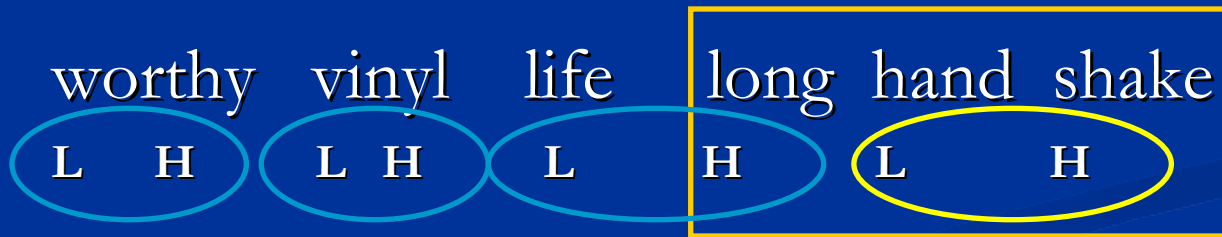
# Condition I: “Pitch”

- F0 alternated between H and L

Monosyllabic context: 



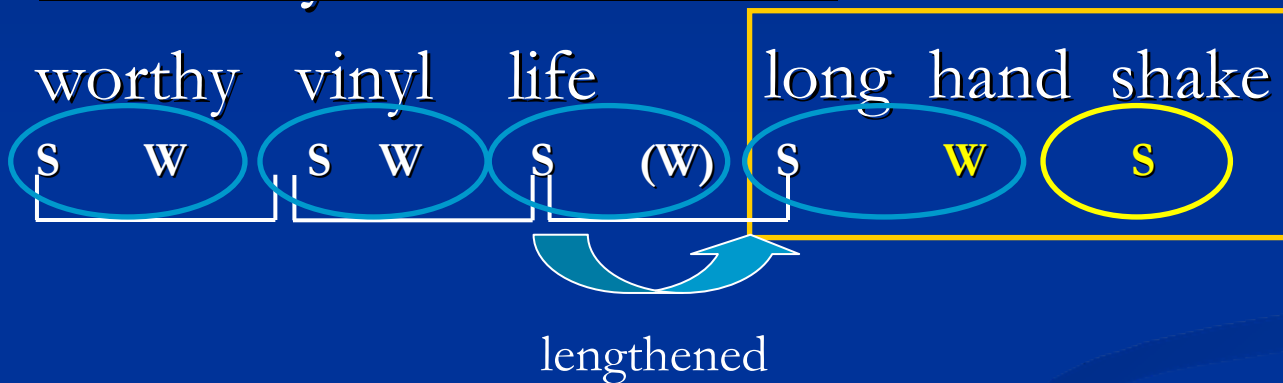
Disyllabic context: 



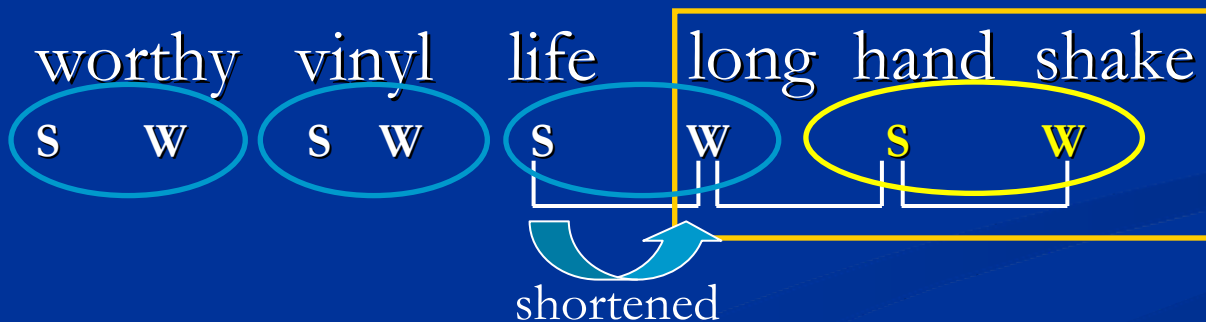
# Condition II: “Duration”

- F0 was flat; interval between syllables 5, 6 varied

Monosyllabic context: 



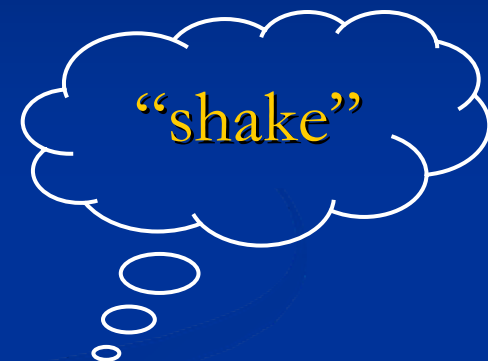
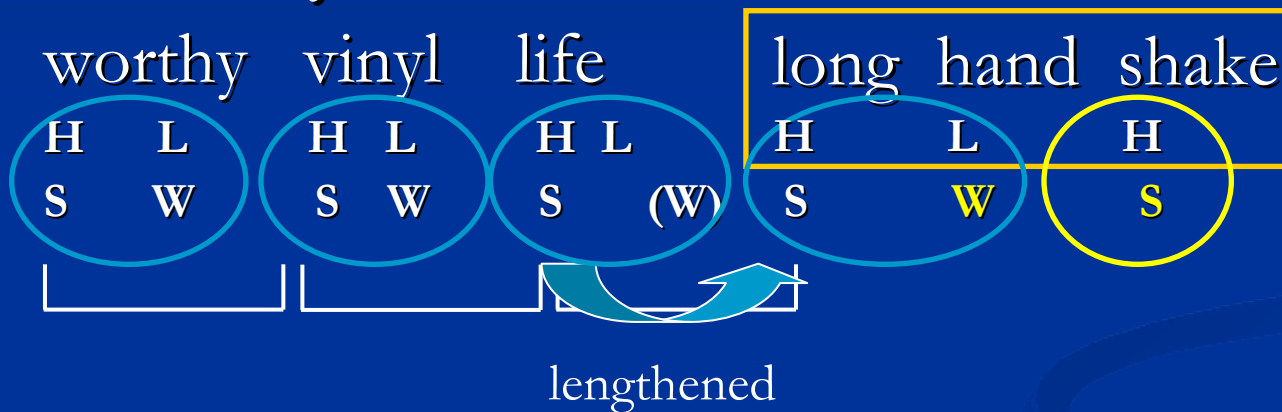
Disyllabic context: 



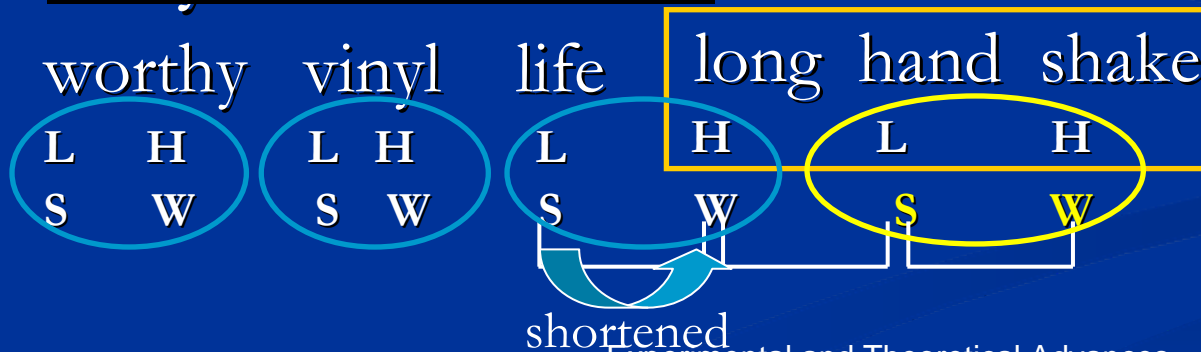
# Condition III: “Pitch + Duration”

- F0 alternated between H and L; interval between syllables 5, 6 varied

## Monosyllabic context:



## Disyllabic context:



# Participants

- One-hundred thirty-eight native speakers of American English attending Ohio State University.
- Assigned to one of the three prosodic conditions.
  - Pitch (n = 57)
  - Duration (n = 40)
  - Pitch + Duration (n = 41)

# Procedure

## ■ Practice

- Participants listened to six filler sequences and wrote down the final word they heard.

## ■ Test

- Participants listened to 100 sequences (20 targets / 80 fillers) and wrote down the final word they heard.
  - 10 targets paired with a disyllabic context
  - 10 targets paired with a monosyllabic context
- Target sequence / context pairing counterbalanced across participants.

# Predictions

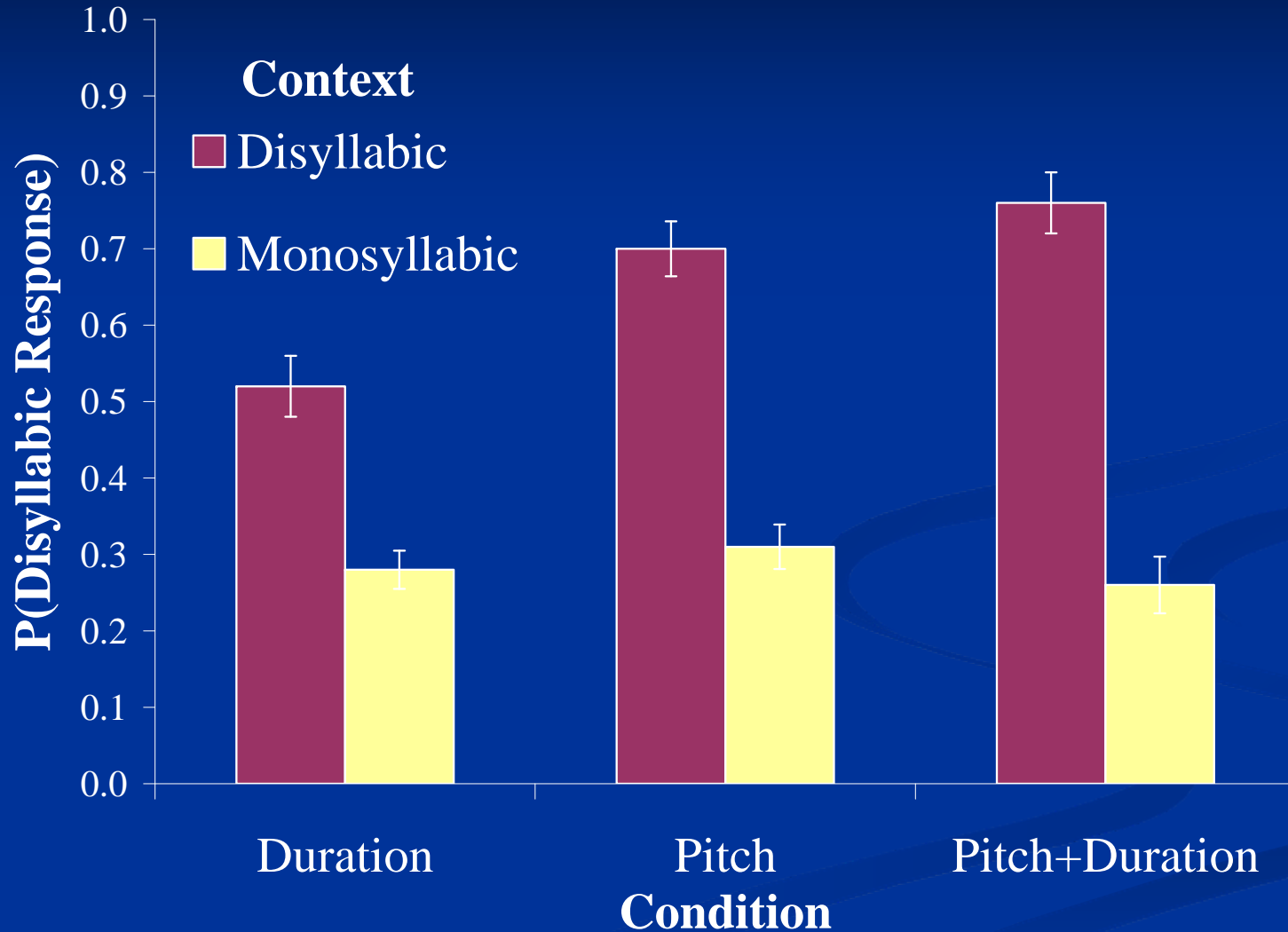
- **Monosyllabic distal contexts** should produce constituent groupings resulting in monosyllabic final word reports:

e.g., *worthy vinyl life longhand shake*

- **Disyllabic distal contexts** should produce constituent groupings resulting in disyllabic final word reports:

e.g., *worthy vinyl lifelong handshake*

# Results



# Follow-up experiments

- Eliminating context syllables 1-4 reduces the size of the distal prosodic effect overall
  - Indicates that the effect is not due entirely to 5<sup>th</sup> syllable
- Replication with cross-modal identity priming
  - Listeners use distal prosody in on-line processing
- Effect holds up to low-pass filtering

# Follow-up experiments, cont'd.

- Better memory for words comprised of syllables which are predicted to be grouped by distal cues
- Distal prosodic cues modulate effects of proximal prosodic cues (e.g., phrasal boundaries)
  - Distal cues can strengthen or wipe out grouping by boundary-related proximal phonetic cues

# Summary

- Distal prosody affected how syllables were perceived as grouped into words, and thus, into prosodic constituents
  - Word boundaries are prosodic boundaries; thus, distal prosody affected presence of prosodic boundaries
- There were more disyllabic responses when prior context favored a disyllabic grouping
- Both pitch and duration were effective cues to structure; combined cues were most effective

# Summary, cont'd.

- Distal prosodic context effects were predicted by principles of general auditory perception
- Identified a new factor – distal prosody – which may influence perception of word boundaries and prosodic boundaries
- Distal prosodic context, not proximal phonetic cues, may be responsible for some cases of constituent perception

# Acknowledgments

- Dr. Devin McAuley, Bowling Green State Univ.
- Dr. Sven Mattys, University of Bristol
- Louis Vinke, Bowling Green State University
- Dr. Stefanie Shattuck-Hufnagel, MIT
- RAP Lab, Bowling Green State University
- Conference organizers

X X X X X X X  
 X X X X X X X

[[*channel*]<sub>F</sub>]<sub>Pwd</sub> [[*dizzy*]<sub>F</sub>]<sub>Pwd</sub> [*foot*]<sub>F</sub> [*note*]<sub>F</sub> [*book*]<sub>F</sub> [*worm*]<sub>F</sub>

X		X		X				
X		X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X

[[*channel*]<sub>F</sub>]<sub>Pwd</sub> [[*dizzy*]<sub>F</sub>]<sub>Pwd</sub> [*foot*]<sub>F</sub> [*note*]<sub>F</sub> [*book*]<sub>F</sub> [*worm*]<sub>F</sub>

H\* L%      H\* L%      H\* L%      H      L      H

X		X		X		X		X
X		X		X		X		X
X	X	X	X	X	(X)	X	X	X (X)

[[[*channel*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[[*dizzy*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[*foot*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[*notebook*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[*worm*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub>

[H\* L<sup>0</sup>%]<sub>IP</sub> [H\* L<sup>0</sup>%]<sub>IP</sub> [H\* L<sup>0</sup>%]<sub>IP</sub> [H\* L<sup>0</sup>%]<sub>IP</sub> [H\* H<sup>0</sup>%]<sub>IP</sub>

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

[[*channel*]<sub>F</sub>]<sub>Pwd</sub> [[*dizzy*]<sub>F</sub>]<sub>Pwd</sub> [*foot*]<sub>F</sub> [*note*]<sub>F</sub> [*book*]<sub>F</sub> [*worm*]<sub>F</sub>

L\* H%

L\* H%

L\*

H

L

H

X		X		X		X	
X		X		X		X	
X	X	X	X	X	X	X	X

[[[*channel*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[[*dizzy*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[[*footnote*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub> [[[*bookworm*]<sub>F</sub>]<sub>Pwd</sub>]<sub>IP</sub>

[L\* H<sup>0</sup>]IP [L\* H<sup>0</sup>]IP [L\* H<sup>0</sup>]IP [L\* H<sup>0</sup>]IP