Prosody: Temporal Structure
Functions of Prosody

• Phrasing (Grouping)
  • When you make hollandaise slowly, it curdles.
  When you make hollandaise, slowly it curdles.

• Prominence (focus)
  • She didn’t earn an A. (she earned a B).
  She didn’t earn an A. (but she got one by cheating).

• How are these functions accomplished by speech system?
  • Temporal structure
  • Intonation
Prosodic Hierarchy

• Speech gestures are organized into larger units, beginning with syllables

• No agreement on exactly how many categories (e.g. intermediate phrase (ip)).

In November, the region’s weather was unusually dry

Prosodic domains above the word: influenced by syntactic structure as well as rhythm

Prosodic domains below the word: influenced by rhythmic/lexical properties

What are the temporal signatures of phrase boundaries are how are they controlled?

What are the temporal signatures of foot structure are how are they controlled?

To keep in mind…

• Given the gestural score for a given syllable, what changes in the score could result in lengthening or shortening of the syllable?

“mad”

• Changes in activation intervals
• Changes in relative timing
Boundary-adjacent lengthening

• Vowels (and consonants) are longer at the end of phrase (or at the beginning of a phrase) than within a phrase (in English and many languages.

• I gave a duck to Doug.
  I gave Doug a duck.

• What kind of boundaries produce lengthening?

• How is the lengthening controlled?
Byrd & Saltzman (1998)

TABLE I. Stimuli sentences for five experimental boundary conditions (Boldface was not present in the stimuli seen by subjects)

<table>
<thead>
<tr>
<th>Boundary condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>none (word-medial)</td>
<td>Poppa begged “mommanini” meanly upon coming.</td>
</tr>
<tr>
<td>word</td>
<td>Poppa-Pikt and Momma-Mimi tapped Coby.</td>
</tr>
<tr>
<td>list</td>
<td>Poppa, Pikt, Momma, Mimi, and Bibi tapped Coby.</td>
</tr>
<tr>
<td>vocative</td>
<td>Quick Momma, Mimi tapped Coby.</td>
</tr>
<tr>
<td>utterance</td>
<td>Poppa picked Momma. Mimi tapped Coby.</td>
</tr>
</tbody>
</table>

Figure 4. The transboundary articulatory interval for three speakers and five experimental boundary conditions. Bars with like shadings group together in post-hoc tests ($p < 0.005$).
Results

• Subjects exhibit one or two levels of lengthening (compared to within-word).

• Relative lengthening of list, vocative, IP varies across speakers.

• How is it accomplished?
  
  • Constrictions approach their goals more slowly (appear to have greater stiffness).

  • Activation intervals are longer.
Gestural model of boundaries: \( \pi \)-gestures

- \( \pi \)-gestures are dynamical cognitive units that represent boundaries.

- Like other gesture they have extent in time (activation intervals) and dynamical parameters that govern how their strength waxes and wanes over time.

- Unlike other gestures, they have no articulators, but rather act vicariously to slow the clock that governs the activation dynamics of all the constriction gestures that fall within its scope.

- Slower activation results in:
  - Gestures are lengthened
  - Gesture onsets are delayed so overlap is reduced.

- Effects are proportional to \( \pi \)-gesture strength.

- Hierarchically higher boundaries will have stronger activations and consequently there will be more lengthening and less overlap than at lower boundaries, thus accounting for the empirical findings that boundary lengthening is cumulative.

Predictions of π-gesture model

• Effect is local.

• All gestures within scope are affected; no gestures are skipped.

• Effect is greatest at the boundary and decreases at a distance from it.

• Lengthening effects all segments, but could interact with type of segment.

• No difference between “final lengthening” and “initial lengthening”

• Categories: no intrinsic differences between phrases of different types (ip) (IP) (e.g. Beckman & Pierrehumbert, 1986); only strength determines effect.

  • Categories could result from multiple modal values of strength or From Boundary Tone Gestures (for IP, but not ip).
Π-gestures as grammatical element

- Tempting to think of slowing as the speaker approaches the end of a phrase as a mechanical consequence of approaching a pause... slow down before stopping.

- But while evidence for boundary-related slowing has been found in several languages (e.g., English, French, Dutch, Greek, German, Spanish, Swedish), it is not universal by any means.

- In the eleven African tone languages from all over Africa in Downing & Rialland (2018):
  - None have final lengthening in declarative IPs
  - Shekgalagari (Botswana) and Tumbuka (Malawi) have penultimate lengthening (common in Bantu).
Reduction in overlap at boundaries

• Much phonetic variation is due to
  
  • variation in the temporal overlap of invariant gestures
  
  • Within a phrase, consonant gestures overlap across word boundaries in English.
  
  • Overlapping gestures jointly determine the output sound, so the acoustic output of a given gesture can vary depending on the gesture that overlaps it.
  
  • Such effects are less likely to occur across boundaries.

• Reduction of activation intervals of gestures within a phrase can also result in truncation: the gesture doesn't reach its goal when speaking rate is fast.
C#C overlap in English vs. Russian

**English**

**Russian**
/t/ “deletion”: /t/ sounds like it is deleted, but it is not

“perfect memory” in three styles

**Boundary**

**Within-phrase**

**Fast**

Release is overlapped by lip gesture and is hidden.

TT gesture is overlapped and truncated.
• “Pack my”
Modeling prosodic variation: “told before”

- Phrase was generated automatically by TaDA/HLsyn, then π-gesture slowing effects were added.
- [d] release emerged automatically with slowing, due to decrease in overlap of Tongue Tip and Lip Gestures.

TaDA:

- No [d] audible
- [d] release

Talker 12

- Apertures: Tongue Tip Lip
- Only 2/38 Transcribers hear boundary

Talker 15

- 11/38 Transcribers hear boundary

Time (s)
Palatalization of /s/

• /s/ within a phrase exhibits apparent change to [ʃ] before [j]
  • “miss it” [mɪʃ]
  • “miss you” [mɪʃ]

• Example
  • “I’m sure I’m gonna miss you”
    slow fast

• What is going on here?
  • We change alveolar fricative to palatoalveolar before [j]?
  • More overlap in faster speech (within a phrase).
“miss. you”

Overlap changes
fricative acoustics

Zsiga (1995)
Figure 20.5. Change in contact patterns over time, subject 1. Electrodes shown in black were activated in at least eight of ten repetitions, those in gray, in seven of ten repetitions.
Lexical and postlexical palatalization

Figure 20.6. Templates from underlying /s/, /ʃ/, and /j/ overlaid on the patterns for s+you and derived /ʃ/ at -3 frames.
Effect of Gestural overlap: Synthesis

SLOW

VELUM WIDE
LIPS STOP
TT ALV FR
TB PALATAL NAR
GLOTTIS WIDE

FAST

VELUM WIDE
LIPS STOP
TT ALV FR
TB PALATAL NAR
GLOTTIS WIDE

PAL NAR VELAR NAR
Place Assimilation: nasal

• Final /n/ is sometimes assimilated to the place of a following labial or dorsal stop:

• “can be”
  [kãenbi] slow vs. [kãembi] fast
Nasal Assimilation: Synthesis

“can be” SLOW

VELUM: WIDE
LIPS: LAB STOP
TT: ALV ST
TB: VEL ST WIDE, PAL NAR
GLOTTIS: WIDE

“can be” FAST

VELUM: WIDE
LIPS: LAB STOP
TT: ALV ST
TB: VEL ST, PAL NAR
GLOTTIS: WIDE

Nasal Assimilation: Synthesis
• syntax: prosodic structure mediates 

• Directly controlled by syntax Cooper & Paccia-Cooper 
  1980, Wagner 2005)

• Prominence — also use of clock-slowing?
  • Coordination of PI gesture
## Prominence and Focus

<table>
<thead>
<tr>
<th></th>
<th>Prompt:</th>
<th>Test sentence:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CF</strong></td>
<td>Is the botanist going to test the fly with the stripes?</td>
<td>No, the botanist will be testing the <em>bee</em> with the stripes.</td>
</tr>
<tr>
<td><strong>NF</strong></td>
<td>What is the botanist going to test?</td>
<td>Oh, the botanist will be testing the <em>bee</em> with the stripes.</td>
</tr>
<tr>
<td><strong>BF</strong></td>
<td>What is the botanist going to do?</td>
<td>Oh, the botanist will be testing the <em>bee</em> with the stripes.</td>
</tr>
<tr>
<td><strong>UA</strong></td>
<td>Is it the zoologist who will be testing the bee with the stripes?</td>
<td>Oh, the <em>botanist</em> will be testing the bee with the stripes.</td>
</tr>
<tr>
<td><strong>UF</strong></td>
<td>Who is going to test the bee with the stripes?</td>
<td>Oh, the <em>botanist</em> will be testing the bee with the stripes.</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Word</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bee</td>
<td>Oh, the botanist will be testing the <strong>bee</strong> with the stripes.</td>
</tr>
<tr>
<td>baby</td>
<td>Oh, his family will be visiting the <strong>baby</strong> with the long hair.</td>
</tr>
<tr>
<td>design</td>
<td>Oh, Ellie’s office will be submitting the <strong>design</strong> for the new building.</td>
</tr>
<tr>
<td>melody</td>
<td>Oh, Jonathan was praising the <strong>melody</strong> from the movie.</td>
</tr>
<tr>
<td>banana</td>
<td>Oh, the baby will be having the <strong>banana</strong> for lunch.</td>
</tr>
<tr>
<td>matinee</td>
<td>Oh, the parents were attending the <strong>matinee</strong> by the sea.</td>
</tr>
<tr>
<td>military</td>
<td>Oh, her cousin will be joining the <strong>military</strong> by the capital.</td>
</tr>
<tr>
<td>humanity</td>
<td>Oh, the historian was applauding the <strong>humanity</strong> of the soldiers.</td>
</tr>
<tr>
<td>salmonella</td>
<td>Oh, the teachers were discussing the <strong>salmonella</strong> for hours.</td>
</tr>
</tbody>
</table>
Results

1. Formation duration (pairwise t-test)

Main effect of Focus

Main effect of Stress

Stress x Length interaction #1

Main effect of Length

Stress x Length interaction #2

All<* except for UA-UF

S1-S2*
S1-S3**
S2-S3 tr. .061

L2-L3**
L2-L4***

All**
All***

L2-L3**
L2-L4***

L1-L3***
L2-L3***
L3-L4**

Length

* * *
2. Release duration (pairwise t-test)

- **Main effect of Focus**
  - CF***
  - NF-BF n.s.
  - NF-UA tr. .057
  - NF-UF* n.s. for others

- **Main effect of Stress**
  - All***

- **Stress x Length interaction #1**
  - All***

- **Main effect of Length**
  - All*** except L2-L4 (.076)

- **Stress x Length interaction #2**
  - All***
  - L2-L3***
  - L3-L4***
  - L2-L4 tr. .051
Modeling Prominence

• $\mu$-gesture
  • Temporal modulation gesture
  • Like $\pi$-gesture, but centered on stressed vowel
Stress and Foot structure

- Organization of syllables into feet

- Possible temporal consequences of foot-structure on rhythm (observed in Germanic languages):
  - Shortening of syllable durations in polysyllabic feet (‘stress-timing’)
  - Greater length of stressed vs. unstressed syllables (vowel reduction)
  - Language differences in rhythm (‘stress-timing’ vs. ‘syllable-timing’)

Foot composed of stressed syllable and following unstressed ones
Polysyllabic shortening

- As syllables are added to a foot, the duration of the syllables decreases.

- Tendency to preserve the duration of the foot.

- This tendency is shown in languages that were traditionally called “stress-timed.” (e.g., Germanic languages)

- Not shown in “syllable-timed” languages in which syllable durations tend to be preserved (e.g., Romance languages)

Interaction of:
- Number of syllables
- Vowel length
- Voicing of coda C
Stress and Polysyllabic shortening


- Stressed syllables are longer than unstressed ones.

![Diagram showing duration of rhyme in stressed and unstressed syllables as a function of the number of syllables in a foot.](image)
Coupled Oscillator Model of Polysyllabic Shotening

- **Hierarchical Coupled Oscillators** (O’Dell & Nieminen, 1999)
  - Harmonically entrained Foot and Syllable oscillators
  - N:1 entrainment: N syllable cycles per 1 foot cycle
  - Hypothesized inter-level asymmetry of coupling strengths can produce polysyllabic shortening in languages that show it:
    - ‘Stress-timed’ languages
      - foot-to-syllable coupling ($\lambda_{FS}$) >> syllable-to-foot ($\lambda_{SF}$)
    - ‘Syllable-timed’ languages
      - foot-to-syllable coupling ($\lambda_{FS}$) << syllable-to-foot ($\lambda_{SF}$)
Polysyllabic shortening Simulation

Saltzman et al., 2008

\[ \lambda_{FS} = 5 \]
\[ \lambda_{SF} = 1 \]
\[ \omega_{0F} = 1 \]
\[ \omega_{0S} = 2 \]

2 syllables per foot

3 syllables per foot

Foot oscillator

Syllable oscillator

Time (s)
Stress Asymmetry

• How can the differential durations of stressed and unstressed syllables be modeled?

• Hypothesize clock slowing gesture ($\mu_T$) that is active at phases of Foot oscillator corresponding to stressed syllables (similar to $\pi$-gesture).

• $\mu_T$ slows clock of Foot and Syllable oscillators (and all constriction gesture) in proportion to its activation level ($a_\mu$).

• Maximum strength of $\mu_T$ gesture will determine the degree or temporal asymmetry between stressed and unstressed syllables.
2 syllables per foot

Foot oscillator

Syllable oscillator

3 syllables per foot

Foot oscillator

Syllable oscillator

Stress Asymmetry Simulation

\[ a_{\mu} \]
Asymmetry in Polysyllabic shortening

- Polysyllabic shortening affects stressed syllables more than unstressed ones.
  - modulate coupling strength asymmetry as a function of phase of the foot oscillator
  - STRESS: \((\lambda_{FS}) > (\lambda_{SF})\)
  - UNSTRESS: \((\lambda_{FS}) < (\lambda_{SF})\)