Coordination of speech gestures and syllable structure
Syllabification

• As Ladefoged notes, English speakers (and speakers of other languages) generally agree about the number of syllables in a word. Some problematic cases:

• (a) "prism, mysticism"
  (b) "peel, seal"
  (c) "hire, fire, hour"
  (d) "mediate, heavier, neolithic"

• How can syllables be defined phonetically?
  • Sonority Peaks
  • Jaw cycles
  • Gestural Coordination
Sonority Peaks

- The number of syllables corresponds to the number of distinct amplitude peaks in the acoustic signal. Why?
- Each peak corresponds to the nucleus of a syllable.
- Sonority principle can account for relatively clear cases.
- It could also explain the disagreements in (1). The "extra" syllable corresponds to a local peak of amplitude that is greater what precedes it, but is not as high as a typical nucleus. E.g,
Problems for Sonority

hidn emz

'hidden aims'

hidnc imz

'hid names'

'sport'

'support'

'bray'

'beret'
Jaw Cycles

• Frame-Content Theory (McNeilage & Davis)

• Speech develops in the infant initially through the oscillation of the jaw.

• Each cycle of the jaw oscillation corresponds to a syllable and is called by M&D ‘the frame.’

• The individual consonants and vowels are “the content” and develop later.

• For adults, the nucleus corresponds to the downward phase of the jaw cycle, the onset and coda correspond to the upward phase.

• In careful speech, jaw behavior differentiates "support" from "sport."

• However, for faster, more casual speech, the "extra" jaw deflection for "support" is absent.
How are gestures glued to one another in time?

- Relative timing of gestures carries information.
- How is appropriate relative timing maintained?
- What is the glue?
- Should make predictions about observed gesture combinations.
Planning intergestural timing
(Saltzman, Nam, Goldstein)

- Each gesture is associated with a planning oscillator, or clock, responsible for triggering that gesture’s activation.

- Relative phase of oscillators (and therefore time of triggering) is controlled by coupling the clocks to one another.
Why clocks and coupling?

- Clocks are nonlinear oscillatory dynamical systems. They exhibit **entrainment:** They **synchronize** with one another.

Demo: Bahraminasab

http://www.youtube.com/watch?v=W1TMZASCR-I
Generality of Entrainment

• Applies to living systems, including humans.  
  http://www.youtube.com/watch?v=4FNolDgNE6o

• Entrainment of clocks within an individual or across individuals

• Coupling doesn’t have to be mechanical.  
  It can be informational (Saltzman, 1995).
Speech entrainment across talkers

- Development of technique for measuring articulator kinematics from two talkers simultaneously.

  - One Carstens EMA, one ND WAVE

  - Talkers (1 M, 1 F) sat 2 m apart facing one another

    - M: “cop top cop top...”
    - F: “top cop top cop...”

Tiede, Kroos, Bundgaard-Nielsen, Gilbert, Attina, Kasiopa, Vatikiotis-Bateson, Best (2010)
BEFORE Entrainment

AFTER Entrainment
Modes of coupling

• Systems of coupled oscillators exhibit distinct modes of synchronization:
  • frequency-locking
  • phase-locking

• These modes have been shown to underlie the coordination of movements of multiple limbs in human action. (e.g., Turvey, 1990; Kelso, 1995).

• The same modes can be used to coordinate speech actions and form an account of syllable structure.
Synchronization modes for limb coordination: phase-locking

- Two relative phase modes (or attractors) are spontaneously available (require no learning) Haken, Kelso & Bunz, 1985
  - 0° (in phase) most stable
  - 180° (anti-phase)

- Oscillation frequency (rate) is a control parameter:
  - Spontaneous transitions to most stable mode (0°) as frequency increases.
  - Fluctuations in phase during transition interval.

Turvey, 1990
Modes & Syllable Structure

• If a basic consonant constriction (C) gesture and a vowel (V) constriction gesture are to be coordinated in a spontaneously available mode, there are just two possibilities:

  • in-phase
    • hypothesized for C-V (onset) simplest, most stable, accessible
  • anti-phase
    • hypothesized for V-C (coda)
Evidence for C-V and V-C modes

Onset C and V gestures begin synchronously \cite{lofqvist1999}; hypothesize that clocks are in-phase.

Coda C begins later than V; hypothesize that clocks are anti-phase.
Gestural Score

“two back”

Audio waveforms and gestures at different time intervals.
Syllable Structure Generalizations

• **Universality:**
  CV syllable is possibly only universal type

• **Combinatorial freedom:**
  • Onsets & rime typically combine relatively freely.
  • Combinations within onset and rime can be more restricted

• **Acquisition:**
  CV acquired earlier by infant than VC

• **Weight:**
  Onsets rarely contribute to weight. Codas frequently do
Universality of CV structure

- All languages have CV syllables, but not all languages have VC structures (e.g. Clements, 1990).

- This can be accounted for by the fact that in-phase is the more accessible, more stable mode.
Combinatorial Freedom

- Combination is free where the coupling mode is maximally accessible without learning (in-phase).

- Combinations are most restricted where learning is required.

<table>
<thead>
<tr>
<th></th>
<th>CV: onset-rime</th>
<th>VC: nucleus-coda</th>
<th>CC: within onset,coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freedom</td>
<td>in-phase</td>
<td>anti-phase</td>
<td>other</td>
</tr>
<tr>
<td>Accessibility</td>
<td>no learning</td>
<td>no learning, but less accessible</td>
<td>particular combinations must be learned</td>
</tr>
</tbody>
</table>
Acquisition of syllable structure
(Nam, Goldstein & Saltzman, in press)

• Infants develop onsets (CV) before codas (VC) in all languages. (e.g. Vihman & Ferguson, 1987; Fikkert, 1994)

• Lag in acquisition of codas is shorter in languages that make more frequent use of VC (Roark & Demuth 2000).

• These facts are predicted by a model of a learning agent that includes both:
  • Greater accessibility in-phase mode
  • Attunement to C<->V phase in the ambient language
Planning model  (Saltzman, Nam, Goldstein)

- Phonological input to planning is a **coupling graph:**
  - **NODES:** Specification of gestures and the associated planning oscillators
  - **EDGES:** coupling functions between pairs of planning oscillators.

- At the beginning of planning process, oscillators are set into motion at random phases.

- Coupling forces specific to graph cause the oscillators to settle at stabilized relative phases (Saltzman & Byrd, 2000).

- Once stabilized, timing oscillators trigger the activation of their associated gesture(s).
Timing and Ig-specific syllabification: CC clusters in onset

• If onset is defined by an in-phase relation between C gesture and V, then all onset C gestures should be synchronous with V (and therefore with each other).

• Multiple constriction gestures in onset cluster (e.g., “spats”).

  • Gestures must be at least partially sequential to afford perceptual recoverability.

• What in the coupling graph identifies them as both in the onset?
Competitive coupling hypothesis
(Browman & Goldstein, 2000)

• Specifications in the coupling graph can compete with one another

• C-V coupling
  • All C gestures in onset coupled in-phase with the V.

• C-C coupling
  • C gestures also coupled sequentially (eccentric)

• Prediction: Observed coordination should reveal the presence of both couplings. As Cs are added to an onset:
  • Rightmost C ($C_n$) should shift later with respect to the vowel.
  • Leftmost C ($C_1$) should shift earlier with respect to the vowel.
Modeling shift with competitive coupling

results of competition

C-V phasing

C_1 shifts left

C_n shifts right

Add an additional coordination (C-C phasing)
Evidence for Rightward shift of $C_n$.

"pea pots" vs "pea spots"
Example: “spat”
Rightward shift of $C_n$ as diagnostic for onset?

- If a consonant sequence is syllabified as part of an onset, then it should exhibit rightward shift.

- Georgian and Tashlhiyt Berber are languages in which words can begin with sequences of 3 obstruents.

- But they differ as in syllabification of such words:
  - Georgian Cs are complex onsets
  - Berber only allows a single C in onset, other Cs constitute nuclei of additional syllables.

- Do Georgian and Berber differ in rightward shift?
Georgian: Rightward shift EMA data (w/Chitoran)

Lag: Target (V) - Target (C_n)

p < .001

2 speakers

/p/ Lip Aperture
/t/ Tongue Tip Constriction Degree
/k/ Tongue Dorsum Constriction Degree
/V/ Tongue Body Constriction Degree

/karebi/ /t'skarebi/ /pt'skaredi/ /riala/ /k'riala/ /t'sk'riala/
Tashlhiyht Berber: Rightward shift EMA data (w/Selkirk)

Lag: Target (V) - Target (Cₙ)  ms.

<table>
<thead>
<tr>
<th></th>
<th>Lag (ms)</th>
<th>2 speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>/m/</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>/t/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/s/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/V/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evidence for complex onsets

• Differences between Georgian and Berber provide tentative support that c-center can be used as diagnostic for complex onset structures.

• Support from other languages:
  • Italian: /sC/ vs. /Cl/ (Hermes, 2013)
  • Moroccan Arabic (Shaw et al. 2010)
  • Romanian (Marin 2012)
C and V gesture valences

• C and V gestures are differentiated by
  • degree of constriction (V is wider)
  • dynamic stiffness (V takes longer to get to target)
  • activation interval (V still active after C released)

• Nature of these differences is such that C and V gestures can be triggered synchronously and still be both be recoverable by listeners (Mattingly, 1981).

• These gestural properties, together with the stability of in-phase coupling gives rise to valence of C and V gestures -- they combine freely with each other in C-V structures.
What gestures can serve as syllable nuclei?
Is there a language in which any segment can be syllabic?


**Examples**

- **Voiced fricative**: /\textipa{tśbGt}/ [ts.b\textipa{Gt}] « you paint »
- **Voiceless fricative**: /\textipa{t-sti}/ [ts.ti] « she chose »
- **Voiced stop**: /\textipa{t-g°ra}/ [tg°.ra] « she took »
- **Voiceless stop**: /\textipa{tk.mi}/ [tk.mi] « she
Figure 1. Audio signal and spectro of one repetition of [sfqqst] by R_R

Figure 2. Audio signal and spectro of one repetition of [tfsxt] by A_R
**Schwa Unstable**

- within one same subject and one same form.

*Acoustic waveform and spectrogram of two repetitions of sfqqst « irritate him » by E_M one with internal schwa (left) and one with final schwa (right).*
This same subject may realize long voiceless sequences with no vowel at all. This is the case for instance for the items \textit{fqqs}, \textit{tfktstt} and \textit{tftXtstt}.

\begin{center}
\begin{tabular}{cccccccc}
\hline
\textit{t} & \textit{f} & \textit{t} & \chi & \textit{t} & \textit{s} & \textit{tt} \\
\hline
\end{tabular}
\end{center}

\textit{Acoustic waveform and spectrogram of one repetition of \textit{[tftXtstt]} « fadeaway » by E._M.}
Evidence for syllabication: Versification

• TB versification distinguishes between heavy and light syllables (Jouad 1983).

• In TB poetry it is common for all the lines of a piece to be sung to the same tune.

• If a tune is comprised of $n$ successive notes, a text with more than $n$ syllables cannot be sung to it.

• (See Dell & Elmedlaoui 2002 and the references therein)
**Sequences sung to a tune**

I present below the parsing of three lines of *Immi nna* “my dear mother” sung by **Iznzarn** (1970)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>im</td>
<td>mi</td>
<td>n</td>
<td>na</td>
<td>kf</td>
<td>rab</td>
<td>biG</td>
<td>ta</td>
<td>sa</td>
<td>wa</td>
<td>la</td>
<td>wu</td>
<td>li</td>
<td></td>
</tr>
<tr>
<td>tz*</td>
<td>da</td>
<td>tz</td>
<td>dm</td>
<td>ta</td>
<td>g°md</td>
<td>Glh</td>
<td>ma</td>
<td>wa</td>
<td>la</td>
<td>sm</td>
<td>mi</td>
<td>di</td>
<td></td>
</tr>
<tr>
<td>ad</td>
<td>da</td>
<td>ln</td>
<td>sa</td>
<td>sm</td>
<td>mid</td>
<td>wat</td>
<td>sm</td>
<td>ms</td>
<td>tr</td>
<td>fl</td>
<td>la</td>
<td>Gi</td>
<td></td>
</tr>
</tbody>
</table>

* This syllable and the following three are emphatic
In-phase C-C syllables in Berber?

• Any segment may appear as nucleus in Tashlhiyt Berber (Dell & Elmedlaoui, 1985).
  • [tu.da] ‘suffice’          [tb.da] ‘begin’

• Expected graphs?
CC vs. CV syllables

Lip Aperture

Tongue Tip Constriction

Time
CC syllables

- C gestures are not in-phase
- If they were, they might not be able to contrast in order
- Alternative graphs: Note abstract syllable oscillator
CC in onset vs. coda: possible coupling graph differences and weight

• Hypothesis: No competitive coupling in coda (for English)

Why?

• Perhaps the weaker anti-phase coupling doesn’t attract the (more remote) as strongly as does the in-phase coupling of onsets.
Onset-Coda asymmetry in weight  
(Nam, 2007)

- **Onset** Cs typically do not contribute to syllable weight.
- **Coda** Cs may or may not depending on the language.

- If weight is related to (syllable) duration, then proposed coupling structures can account for the difference between onset and coda consonants in weight.
- With synchronous onset coupling, effect of rightward shift is that adding a C to onset does not increase syllable duration by the duration of the C (more like one-half the C duration).

Languages in which coda Cs do not bear weight are predicted to show competitive coda coupling (e.g., Malayalam, Broselow et al. (1997))
C-Center in coda?
(Marin & Pouplier, 2010)

• Predictions of coupling asymmetry model:

Onsets:
• c-center stability
• rightward shift

Codas:
• left edge stability
• no leftward shift
Results: shift

- /s-stop/ support asymmetry
- clusters with liquids do not (can be explained by a graph topology that explicitly incorporates multiple gestures for /l/)

\[
\begin{align*}
\text{Shift 67 ms} & \quad \text{s} \quad \text{scab} \\
\text{Shift 9 ms} & \quad \text{k} \quad \text{cab} \\
\text{bask} & \quad \text{k} \quad \text{s} \\
\text{bass} & \quad \text{s} \\
\end{align*}
\]
Timing stability in onsets vs. codas

Timing between C gestures is more stable in onset clusters than in coda clusters (Byrd, 1996).
Simulation
(Nam, 2007)

- Hypothesis: Loop topology of onsets adds stability (multiple paths) compared to chain.
- Add noise to simulations
  - Noise source: $\xi_i(t) = \text{Gaussian, zero mean, unit variance}$
  - st.dev. of noise (“strength”), $\beta$, varied across conditions
- Result: Greater steady-state relative phase stability (lower standard deviation, $\sigma_{ss}$) for clusters in onsets than codas

![Graph showing standard deviation of C-C phase (radians) vs. standard of noise for onsets and codas.]

<table>
<thead>
<tr>
<th>std. of noise</th>
<th>Onsets</th>
<th>Codas</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>0.25</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>0.45</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>0.65</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>0.85</td>
<td>0.65</td>
<td>0.85</td>
</tr>
</tbody>
</table>