Dynamical models of prosody
What can we learn about prosody using dynamical models?

1. Some commonly occurring ("unmarked") phonological structures may be understood as attractors of speech production dynamical systems.

2. Principled relation between qualitative phonological forms and quantitative observables and parameters: how do distinct qualitative phonological forms result from changes in continuous parameter values?
   
   - speech errors
   - sound change
   - prosodic typology (e.g., quantity sensitivity)
   - rhythmic typology
   - Interactions among levels in the prosodic hierarchy
   - acquisition of novel prosodic phonological forms.
Coupled oscillatory dynamical systems

- Coupled oscillators exhibit entrainment: They synchronize with one another.

- Entrainment applies to living systems, including humans.

- Entrainment of oscillators within an individual or across individuals.

- Coupling doesn’t have to be mechanical.

- It can be informational (Saltzman, 1995).
Modes of synchronization

- Systems of coupled oscillators exhibit qualitatively distinct modes of synchronization (attractor states):
  - phase-locking
  - frequency-locking
- These modes have been shown to underlie the coordination of movements of multiple limbs in human action. (e.g., Turvey, 1990; Kelso, 1995).
- Applying these modes to the coordination of speech production actions at multiple levels reveals attractors that can underlie unmarked phonological/prosodic structures.
- The relative stability of those modes can be influenced by the value of the system’s control parameters. As the values change, the state of the system can shift into a different mode.
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)
- frequency (rate) is a control parameter:
  - Spontaneous transitions to most stable mode (0°) as frequency increases.
Coupled oscillator model of syllable structure
(Saltzman, Nam, Goldstein)

- Each articulatory 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) can be controlled by coupling the clocks to one another.
- Two possible stable modes:
  - in-phase hypothesized for C-V (onset relation) simplest, most stable
  - predicts C-V onset synchrony
  - anti-phase hypothesized for V-C (coda relation)
Universality of CV syllables

• The fact that in-phase is the more accessible, more stable mode contributes to an account of syllable structure markedness and typology:
  • All languages have CV syllables.
  • Not all languages have (C)VC syllables.
Coupling Graph Model of Syllable Structure

- Syllable structure is modeled by topology of coupling graph:
  - NODES: Specification of gestures and the associated planning oscillators.
  - EDGES: Coupling functions that specify preferred mode between pairs of planning oscillators.

- Relatively free combination of onsets and rimes.
- Shorter planning time for CV syllables.
- Coordination in CCV vs. VCC clusters.
- Preference for onsets to be weightless.
- Greater stochastic variability in relative timing of coda clusters than onsets.
- Coordination of tone and pitch accent gestures with CV gestures.
Qualitative changes in phonological structure
Qualitative changes can be caused by continuous changes in conditions: speech errors

- Grammatical modes of a system (as represented in the coupling graph) coexist with intrinsic attractor modes.
- The relative stability of modes can be influenced by the value of the system’s control parameters.
  - As the control parameter values change, the state of the system can shift into a different mode.
- Speaking conditions (e.g. rate, prosody, phonological & morphological context) can function as control parameters:
  - Can move the coordination pattern out of the grammatical mode into a new mode and result in a speech error.
Speech Errors

- Examine gestures during repetitive tasks that elicit speech errors.
  - EMA (Pouplier, Tiede, Mooshammer)
  - real-time MRI (Proctor, Lammert, Hagedorn)

“cop top cop top...”
Gesture Intrusions

- “Extra” copy of a gesture appears at an inappropriate temporal location.
- Most common error observed.
- Usually co-produced with the gesture that is appropriate at that time (reductions occur but much less frequently).
- Perceptual consequence (transcription) varies depending on several factors, e.g. magnitude of intruding gesture (Pouplier & Goldstein, 2006).
  - Examples with high magnitudes are perceived as “errors,” i.e. differently what what was intended.
Intrusion error

Overlaid frames
Why do intrusions occur?

- Under the conditions of repetition, individual articulators (LL, TT, TB, jaw) function as nonlinear oscillators that achieve the repeating constriction task.

- The nonlinear oscillators entrain.

- Under control of rate (and/or other parameters), they can exhibit a shift to a different mode (which no longer achieves the linguistic task).

- What are the relevant modes?
Synchronization modes for multi-frequency limb coordination: frequency-locking

- When performing oscillatory motions of multiple body parts
  - Simple rhythms (e.g., 1:1, 2:1) can be performed without learning or practice.
  - Complex multi-frequency rhythms (e.g., 4:3, 5:2) must be learned.
- Spontaneous transitions are observed from complex to simpler modes as a function of frequency.
- Low-order modes are stable attractor states.

Skilled drummers performed bimanual tapping with a 5:2 frequency-locking and gradually increase movement frequency.

“cop top” intrusion errors

- **TT** and **Lips** are frequency-locked in 1:2 relation.
- As rate increases, TT intrusion error can be viewed as a fluctuation enroute to transition to a simpler mode of frequency-locking (1:1) between TT and Lip constrictors.
Persistent qualitative changes: sound changes

• We might expect certain qualitative shifts might result in long-term changes in the phonological organization of a particular language.

• persistence of the control parameter that gives rise to it?
Mode shift in Western Andalusian Spanish: Parrell (2012)

- Many dialects of Spanish, including Western Andalusian Spanish (WAS), show aspiration of /s/ in coda position:

/kasta/ → [kahta] (“caste”)

Analysis: a reduction in the oral constriction gesture for /s/ while maintaining the glottal spreading gesture for voicelessness (e.g. Romero 1995).

- WAS shows postaspiration of voiceless stops following coda /s/ (Torreira 2007a).

/kasta/ → [kata]\ a]

- Production of /s/ is variable:

/kasta/ → [kahta], [kata] [kat\ a]
Coupling graph analysis

[kahta]

V  V

GLC wide → TT clo

wide

closure

preaspiration closure VOT

[katha]

V  V

GLC wide → TT clo

wide

closure

VOT

cf. Torreira (2007a, 2007b)
Mode shift hypothesis

- Shift in the relative phase of the glottal abduction and stop closure planning oscillators observed in WAS **can be induced** by increasing rate:
  - The grammatically stable mode (**eccentric** relative phase) will be destabilized under rate increase and the system will fall into the **intrinsically** accessible **in-phase** attractor.
Rate Experiment

• Gradually increasing speaking rate should trigger mode shift.

• Experimental predictions:

  1. As rate increases, speakers will shift from preaspiration + short VOT → no preaspiration and long VOT

  2. Speakers will never switch in the reverse direction

  3. Temporal variability will be greatest during the phase transition, and greater before the phase switch than after, because in-phase mode is most stable (e.g., Kelso 1984).
Results

- As predicted, speakers switched from short to long VOT (28 trials) but not from long to short (0 trials)

\[
\begin{array}{c|cc}
\text{INITIAL VOT} & \text{SHORT} & \text{LONG} \\
\hline
\text{SHORT} & 18 \text{ (SS)} & 28 \text{ (SL)} \\
\text{LONG} & 0 \text{ (LS)} & 33 \text{ (LL)} \\
\end{array}
\]

\[\chi^2 = 16.7234, p < 0.0001\]
Results

• measure of variability:
  Average deviation from mean aspiration ratio

\[
\frac{VOT}{VOT + Preaspiration}
\]

DURING SHIFT > BEFORE SHIFT > AFTER SHIFT

F(2, 831) = 49.4903, p < 0.0001

As predicted, variability in production was greatest during the phase transition, and higher before (less stable coordination) than after.
The prosodic hierarchy
Prosodic hierarchy as ensemble of coupled multi-frequency oscillators

- Prosodic hierarchy can be conceptualized as a network of oscillators of different frequency: Higher levels are associated with lower frequencies.

\[ R^2 = 0.26, \quad F = 10.85, \quad p < 0.005 \]

\[ R^2 = 0.20, \quad F = 7.71, \quad p < 0.01 \]

\[ \omega_{Phr} = 1 \]
\[ \omega_{Ft} = 2 \omega_{Phr} \]
\[ \omega_{\sigma} = 2 \omega_{Ft} \]
\[ \omega_{C_1V, C_2} = \omega_{\sigma} \]

\[ \text{Phr} \]
\[ \text{Ft} \]
\[ \text{C} \]
\[ \text{V} \]
\[ \text{sp} \]
\[ \text{a} \]

Tilsen (2009)
Preference for simple integer modes of frequency locking

• Basis for strict-layering (Selkirk, 1986)
  • units of lower levels of prosodic hierarchy are exhaustively contained within units of higher levels
Coupling between Foot and Syllable oscillators

- How are syllables organized into stress feet?
- One common pattern quantity-sensitivity:
  - two light syllables form a foot (CV CV)
  - one heavy syllable can form a foot (CVC) or (CVV)
- Number of syllables per foot can be treated as $F:\sigma$ frequency-locking mode
  - 1:1 for heavy $\sigma$, 1:2 for light $\sigma$
- Goldsmith (unpub) proposed that this difference in locking mode can emerge as a bifurcation in the behavior of a coupled F-$\sigma$ system as a function of the frequency $\omega_\sigma$ of the syllable oscillator (1/cycle duration):
  - heavy $\sigma$: cycle duration is long and $\omega_\sigma$ is low $\rightarrow$ 1:1 emerges
  - light $\sigma$: cycle duration is short and $\omega_\sigma$ is high $\rightarrow$ 1:2 emerges
Goldsmith’s model: coupled phase oscillators

\[ \dot{\theta}_\sigma = \omega_\sigma - \alpha \sin(\theta_\sigma - \theta_F) - (1 - \alpha) \sin(\theta_\sigma - 2\theta_F) \]
\[ \dot{\theta}_F = \omega_F + \alpha \sin(\theta_\sigma - \theta_F) + (1 - \alpha) \sin(\theta_\sigma - 2\theta_F) \]

- \( \alpha \) is the control parameter specifying the nature of the coupling
- Observe the behavior of the system when coupled (\( \alpha = .5 \)) vs. uncoupled (\( \alpha = 0 \)) for different values of \( \omega_\sigma \)
Results

\( \alpha = 0 \) (uncoupled)

- Heavy (\( \omega_\sigma = 1.45 \), \( \omega_F = 1 \))
- Light (\( \omega_\sigma = 1.75 \), \( \omega_F = 1 \))

\( \alpha = 0.5 \)

- Heavy (\( \omega_\sigma = 1.45 \), \( \omega_F = 1 \))
- Light (\( \omega_\sigma = 1.75 \), \( \omega_F = 1 \))
Implications

• Bifurcation between 1:1 mode and 1:2 mode as a function of a changes to one of the system’s control parameters $\omega_\sigma$ (critical value of $\omega_\sigma = 1.55$)

1. Qualitative difference in phonological form emerges from quantitative parameter values.

2. Different foot types (H) (LL) correspond to distinct modes of a coupled oscillatory system.

• What about non-quantity sensitive systems?

• setting $\alpha < .5$ will push critical value lower, so effectively all feet will be bisyllabic:

$$\alpha = 0, \quad \omega_\sigma = 1.33, \quad \omega_F = 1$$

$$\alpha = .3$$

3. Setting the value of $\alpha$ creates distinct prosodic grammars.
When $\omega_F \sim \omega_\sigma$ and 1:2 coupling is engaged, the output periods must accommodate:

- period of $F >> 1/\omega_F$
- period of $\sigma << 1/\omega_\sigma$

O’Dell and Nieminen (1999) model: amount of accommodation shown by the $F$ or $\sigma$ oscillators can be asymmetric

- a control parameter ($r$) governs the relative strength of the coupling: $F \rightarrow \sigma$ vs. $\sigma \rightarrow F$

- Changing $r$ produces patterns like traditional “stress-timing” vs “syllable-timing” as the number or syllables per foot ($n$) increases

- “stress-timing”: period of $\sigma$ shortens more than period of $F$ lengthens
- “syllable-timing” period of $\sigma$ shortens less than period of $F$ lengthens
O’Dell and Nieminen’s (1999) model

- Eriksson (1991) fit linear models showing how inter-stress interval \((i)\) is a linear function of the number \((n)\) of \(\sigma\) per \(F\).

\[
I = a + bn
\]

\[
\theta_F = \omega_F + H(\theta_\sigma - n\theta_F)
\]

\[
\theta_\sigma = \omega_\sigma - rH(\theta_\sigma - n\theta_F)
\]

\[
I(n) = \frac{1}{r\omega_F + \omega_\sigma} n
\]

\[
r = \frac{a}{b}
\]

<table>
<thead>
<tr>
<th>Language</th>
<th>(I = )</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>(201 + 102n)</td>
<td>0.996</td>
</tr>
<tr>
<td>Thai</td>
<td>(220 + 97n)</td>
<td>0.973</td>
</tr>
<tr>
<td>Spanish</td>
<td>(76 + 119n)</td>
<td>0.997</td>
</tr>
<tr>
<td>Greek</td>
<td>(107 + 104n)</td>
<td>1.000</td>
</tr>
<tr>
<td>Italian</td>
<td>(110 + 105n)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 1. Linear regression equations and correlation coefficients \((r)\) for five languages using Dauer’s [2] data.
Implication

- Dynamical coupling model can account for both qualitative grammars and quantitative differences between languages.
Durational Asymmetry of stressed vs. unstressed $\sigma$

- In many languages, stressed syllables are longer than unstressed.
- O’Dell and Nieminen do not model the duration of individual $\sigma$, only the effect of coupling on F durations.
- Saltzman, Nam, Goldstein & Krivokapic (2008) developed a similar coupling approach but also attempted to model within-F stress asymmetries.
Polysyllabic shortening Simulation

$\lambda_{F\sigma} = 5$

$\lambda_{\sigma F} = 1$

$\omega_{0F} = 1$

$\omega_{0S} = 2$

2 syllables per foot

3 syllables per foot

Time (s)
Stress Asymmetry

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

• Hypothesize clock slowing gesture (µ_T) that is active at phases of Foot oscillator corresponding to stressed syllables.

• µ_T slows clock of Foot and Syllable oscillators (and all constriction gesture) in proportion to its activation level (a_µ).

• Maximum strength of µ_T gesture will determine the degree of temporal asymmetry between stressed and unstressed syllables.
Stress Asymmetry

- 2 syllables per foot
- 3 syllables per foot

Foot oscillator
Syllable oscillator

Time

Tuesday, October 2, 12
Beyond the foot: foot-phase interactions

Several studies have shown that lengthening due to prosodic boundaries ("final-lengthening") affects stressed syllables, but not unstressed ones, even when they are remote from the boundary (English: Turk & Shattuck-Hufnagel, 2007; Byrd & Riggs, 2011; Rusaw, 2011; Greek: Katsika, 2011)

Rusaw (2011)
Rusaw’s (2011) model

- 3 neurally-inspired central pattern generator (CPG) oscillators with periods corresponding phrases, feet (stress), and syllables.

- Integrate-and-fire nodes corresponding to each oscillator can be connected to each other with inhibitory or excitatory projections

- inhibition lengthens the period of the inhibited oscillator.

- excitation shortens the period of the excited oscillator.

- cf. π-gesture model (Byrd & Saltzman, 2003) and μ-gesture model (Saltzman et al., 2008)
Modeling foot-phrase interactions

- Inhibitory projection from PHASE to SYLL → boundary lengthening
- Inhibitory projection from STRESS to SYLL → stress lengthening
- Excitatory projection from PHASE to STRESS → interaction
Dynamics of nuclear accent learning  (Nava, 2010)

- **English:** Nuclear accent can be retracted to pre-final position in constructions like broad focus unaccusatives:
  
  A *window* closed.

- **Spanish:** Pre-final nuclear accent for this context is not possible.
Model based on Gafos & Benus (2006)

- Dynamical system that provides multiple possible attractor states along some continuous state variable, $x$ (order parameter).

$$\dot{x} = -\frac{dV}{dx} \quad V(x) = -Rx - \frac{x^2}{2} + \frac{x^4}{4}$$

- Control parameter $R$ determines tilt of potential.

- If order parameter is the Relative Prominence of the final and pre-final constituent, then English can be modeled with $R=0$, but Spanish with $R=1.5$. 

Tilted Anharmonic Oscillator (Tuller et al., 1994)
Nuclear Accent in L2

• Nava found that L1Spanish-L2English speakers of intermediate ability failed to show the flexibility of native English speakers.

• 35 Native, 47 L2 speakers producing 8 unaccusative sentences.

• Durations
  • $V_{\text{final}}$ = stressed syllable of sentence final word
  • $V_{\text{prefinal}}$ = stressed syllable of lexical word preceding final

• Probability Density Distribution of Relative Prominence indexed by $\log(\text{duration}(V_{\text{final}}/V_{\text{prefinal}}))$, fit with two gaussians (GMM model).
Learning as parameter dynamics

• Learning flexibility in nuclear accent placement can be modeled as a change in the value of R, which results in a qualitatively distinct grammar.

• Nava showed that

  • Experienced L2 speakers develop values of R close to 0.
  
  • The change in the value of R was found to be (quantitatively) a function of the asymmetry in stress duration ($\mu_T$) and $(\lambda_{FS}) > (\lambda_{SF})$, as measured in a separate rhythmic task.

  • Changes in foot-syllable coupling precede changes in the accent attractor layout.