Gestural coordination in the living lexicon of spoken words

UCL – Speech Science Forum
February 11, 2021

Jason A. Shaw
Yale University
Last time (July 16th, 2012)

• **Dynamic invariance** in the phonetic expression of syllable structure


• Abstract phonological structure conditions non-arbitrary variation in the kinematics.
Dynamic invariance: variation in the kinematics follows from noisy actuation of coordinated gestures
1) **Dynamic invariance**: still a good idea!
   – Gestural basis for complex segments (Russian, English)

2) Gestural reorganization **conditioned by linguistic context** (language-specific)
   – Gesture deletion triggers re-organization of gestural coordination (Japanese)
   – Morpho-syntax conditions re-organization of gestural coordination (Mandarin)
   – Tone exogenesis with (Mandarin) and without (diaspora Tibetan) re-organization of segmental gestures

3) Living lexicon: word-specific phonetics
   – **Lexical absorption**: words take on the phonetic detail of the prosodic environments in which they are typically produced (Mandarin)
   – **Lexical persistence**: phonetic resistance to structurally-conditioned pitch accent reduction (Japanese).
Direct actuation
(lexical persistence)

Gesture-based lexicon → Contextual reorganization of gestures → ε → kinematics

Feedback loop
(lexical absorption)
1) **Dynamic invariance:** still a good idea!
   – Gestural basis for complex segments (Russian, English)

2) Gestural coordination is **conditioned by linguistic context**
   – Gesture deletion triggers re-organization of gestural coordination (Japanese)
   – Re-organization of gestural coordination precipitates tone loss (Mandarin)
   – Tone loss proceeds without gestural re-organization (diaspora Tibetan)

3) **Living lexicon:** word-specific phonetics
   – **Lexical absorption:** words take on the phonetic detail of the prosodic environments in which they are typically produced (Mandarin)
   – **Lexical persistence:** phonetic resistance to structurally-conditioned pitch accent reduction (Japanese).
Gestural basis for complex segments

collaborators

Sejin Oh, PhD Candidate
CUNY/Haskins

Karthik Durvasula,
Michigan State University

Alexei Kochetov,
University of Toronto
Segment sequences vs. Complex segments

• Descriptively, we recognize **segment sequences** as **distinct from complex segments**:
  – segment sequences: /pj/, /kw/, /kp/, /ps/
  – complex segments: /pʲ/, /kʷ/, /k̞p/, /ps/

• What is the basis for this structural distinction?
Phonological diagnostics for complex segmenthood

• **Contrast:** in rare cases, languages contrast complex segments and segment sequences:

  e.g., Russian Cʲ vs. Cj (near) minimal pairs
  
  a) /lʲut/ ‘fierce’ /ljut/ ‘pour (3p pl)’
  b) /dʲatʲel/ 'woodpecker' /djakon/ 'deacon'
  c) /pʲok/ ‘bake (3ps pst)’ /pjot/ ‘drink (3ps)’
  d) /bʲust/ ‘bust’ /bjut/ ‘beat (3p pl)’

• Morpho-phonological patterns, segment distribution, language games
Phonetic diagnostics for complex segmenthood

• At least in cases of contrast, there must be phonetic differences, but...

• Complex segments are not systematically shorter in phonetic duration than gesturally matched segment sequences (Gouskova & Stanton 2019, c.f., Trubetzkoy 1939)

• We pursue the hypothesis that there is a gestural basis to the distinction with kinematic consequences:

  **HYPOTHESIS** (Shaw, Durvasula, Kochetov, 2019)

  $H$: complex segments involve gestures coordinated according to onset landmarks
Key assumptions ($A_1, A_2$)

$A_1$: **Gestures** are forces that drive articulators to task goals over time (e.g., Browman & Goldstein 1986)

$A_2$: A gesture can be decomposed into a series of states or **landmarks** (Gafos 2002)

Gestures:

- **onset**
- **target**
- **release**
- **offset**

Time:

- **G**
Key assumptions ($A_3$, $A_4$)

$A_3$: **Coordination** relations between gestures make reference to **gesture landmarks**: e.g., the *onset* of $G_2$ is coordinated with the *offset* of $G_1$ (e.g., Gafos 2002)

$A_4$: There may be a **consistent +/- lag** between coordinated landmarks (e.g., Shaw & Gafos 2015)

Segment sequence

<table>
<thead>
<tr>
<th>$G_1$</th>
<th>offset</th>
</tr>
</thead>
</table>

Segment sequence with **negative lag**

<table>
<thead>
<tr>
<th>$G_1$</th>
<th>offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_2$</td>
<td>onset</td>
</tr>
</tbody>
</table>

- $lag = -$
Lag can cause surface ambiguity

Segment sequence – no lag

Complex segment – no lag

Segment sequence – negative lag

Complex segment – positive lag

similar patterns of gesture overlap can derive from different coordination relations
Stochastic models of coordination: approach
(following Shaw & Gafos 2010, 2015; Gafos et al. 2014; Shaw et al. 2011)

**Guiding principle:** phonetic variation derives from noisy actuation of discrete gestures and coordination relations between them.

- Define coordination relations as statistical dependencies between gesture landmarks
- Simulations:
  - **Random variation:** kinematics as noisy actuation of dynamics
  - **Controlled variation:** introduce systematic variation in one phonetic parameter to observe how other phonetic parameters vary.
- Identify differences in **structure-specific covariation** across competing hypotheses.
Random variation: each landmark simulated with noise

Segment sequence, e.g., [pj]

\[ G_1^{\text{Onset}} = G_1^{\text{Offset}} - k^{\text{dur}} + \epsilon \]

\[ G_1^{\text{Offset}} = \mathcal{N}(\mu, \sigma^2) \]

\[ G_2^{\text{Onset}} = G_1^{\text{Offset}} - k^{\text{Lag}} + \epsilon \]

Complex segment, e.g., [p\text{j}]

\[ G_1^{\text{Onset}} = G_1^{\text{Offset}} - k^{\text{dur}} + \epsilon \]

\[ G_1^{\text{Offset}} = \mathcal{N}(\mu, \sigma^2) \]

\[ G_2^{\text{Onset}} = G_1^{\text{Onset}} + k^{\text{Lag}} + \epsilon \]
Controlled variation:

$G_1$ duration varied systematically

Segment sequence, e.g., [pj]

Complex segment, e.g., [pj]

$G_1^{\text{Onset}} = G_1^{\text{Offset}} - k_{\text{dur}} + \epsilon$

$G_1^{\text{Onset}} = G_1^{\text{Offset}} = N(\mu, \sigma^2)$

$G_2^{\text{Onset}} = G_1^{\text{Offset}} - k_{\text{Lag}} + \epsilon$

$G_2^{\text{Onset}} = G_1^{\text{Onset}} + k_{\text{Lag}} + \epsilon$

Varied from 200:250 ms in 1 ms steps

$G_1^{\text{Onset}} = G_1^{\text{Offset}} - k_{\text{Lag}} + \epsilon$

$G_1^{\text{Onset}} = G_1^{\text{Offset}} = N(\mu, \sigma^2)$
Coordination relations constrain phonetic covariation

**Key simulation result:** How variation in $G_1$ duration influences onset-to-onset lag depends on coordination relations.

- Onset-to-onset
- $G_1$ duration
- $G_2$
- Segment sequence
- Complex segment

**Simulation results**
- Positive correlation
- No effect of $G_1$ on onset-to-onset lag

17
Empirical tests

Fleshpoint tracking using EMA & X-Ray Microbeam

(1) Russian palatalized labial vs. control sequence
- Subset of EMMA data from Kochetov (2006)
- 3 female speakers
- /pʲ/ & /br/ sequences
- 2 items per sequence; 4-5 reps

(2) English /bj/ sequences
- Wisconsin X-Ray Microbeam (Westbury 1994)
- 20 speakers, 1 rep per speaker
- Task 33: “beautiful” in word list

(3) Russian vs. English
- New NDI Wave 3D EMA data
- 8 speakers (4 Russian), 20-30 repetitions per item
- Russian: /bʲ/, /pʲ/, /mʲ/, /vʲ/ items in carrier phrase
- English: /bj/, /pj/, /mj/, /vj/ items in carrier phrase
Data measurement

**Gestures** parsed according to primary articulator: tongue blade for [j]; tongue tip for [r] (rhotic trill); lip aperture for [m], [p], [b], [v]

Landmarks: **Onset, Target, Release, Offset** determined by 20% threshold of peak velocity in Mview (Tiede 2005)

**Dependent measures**

\[
G_1 \text{ duration} = \text{Offset (} G_1 \text{)} - \text{Onset (} G_1 \text{)}
\]

\[
G_2
\]

\[
onset-to-onset = \text{Onset (} G_2 \text{)} - \text{Onset (} G_1 \text{)}
\]
Representative token (English data)

“It’s a butte perhaps”

\[
\begin{align*}
&\text{s} & & \text{ə} & & \text{b} & & \text{j} & & \text{u} \\
&\text{onset} & & \text{b} & & \text{offset} \\
&\text{onset} & & \text{j} \\
\end{align*}
\]
Predictions

1. **Segment sequence timing (all English data and Russian /br/)**: the lag between the onsets of gestures increases with the duration of the first gesture.

2. **Complex segment timing (Russian /pʲ/)**: the lag between the onsets of gestures is not affected by the duration of the gestures.

![Diagram](image)
Russian data

(Kochetov 2006)

G1 duration

Onset-to-onset

G1

G2

onset-to-onset

3 speakers; 4-5 reps per item

/brat#p¹atava/ ‘брат пятого’

/brat#padaja/ ‘брат падая’

/tat#p¹api/ ‘тат пяпы’

/ta#p¹api/ ‘та пяпы’
English control data

(X-ray Microbeam)

20 speakers (1 rep) “beautiful”
New EMA experiment

- No main effect of language on onset-to-onset lag.
- Strong interaction between $G_1$ duration and language

Table 1: LME Model comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 + (1</td>
<td>subject)+(1</td>
<td>item)$</td>
<td>4</td>
<td>10270.7</td>
<td>-5131.35</td>
</tr>
<tr>
<td>$1 + G_1$ duration + (1</td>
<td>subject)+(1</td>
<td>item)$</td>
<td>5</td>
<td>10159.6</td>
<td>-5074.8</td>
</tr>
<tr>
<td>$1 + G_1$ duration + language + (1</td>
<td>subject)+(1</td>
<td>item)$</td>
<td>6</td>
<td>10161.6</td>
<td>-5074.8</td>
</tr>
<tr>
<td>$1 + G_1$ duration * language + (1</td>
<td>subject)+(1</td>
<td>item)$</td>
<td>7</td>
<td>10076.9</td>
<td>-5031.5</td>
</tr>
</tbody>
</table>

Onset-to-onset lag (~)
Discussion: gestural basis of complex segments

• Predictions borne out:
  – English labial-palatal gestures timed as segment sequences
  – Russian labial-palatal gestures timed as complex segments

• Phonologically relevant dynamics can be diagnosed in the kinematics because of how coordination relations structure variability (Shaw et al. 2011; see also Oh 2020 on Korean coda nasals)

• Consistent with view of the lexicon as consisting of discrete gestures and coordination relations between them.
Future directions

• We focused here on underlyingly palatalized consonants of Russian, but consonant-glide sequences are also described as “palatal” (Timberlake 1984) while “plain” consonants are velarized/uvularized (Roon et al. 2019), e.g.: 
  \(/\text{p}^\text{jot}/ \text{‘drink (3ps)’} \rightarrow [\text{p}^\text{j}]\text{jot}\)

• Do underlying plain (velarized/uvularized) consonants also show gestural timing characteristic of complex segments? (Oh et al., 2020, in prep)

• Gestural basis of complex segments may generalize to other cases, including those not traditionally thought of as “complex”:
  \(\rightarrow\) pre-nasalized stops, etc., but also aspirated stops, nasals, ....
Dynamic invariance: variation in the kinematics follows from noisy actuation of gestures
This talk

1) **Dynamic invariance**: still a good idea!
   - Gestural basis for complex segments (Russian, English)

2) **Gestural reorganization** *conditioned by linguistic context* (language-specific)
   - Gesture deletion triggers re-organization of gestural coordination (Japanese)
   - Morpho-syntax conditions re-organization of gestural coordination (Mandarin)
   - Tone exogenesis with (Mandarin) and without (diaspora Tibetan) re-organization of segmental gestures

3) **Living lexicon**: word-specific phonetics
   - **Lexical absorption**: words take on the phonetic detail of the prosodic environments in which they are typically produced (Mandarin)
   - **Lexical persistence**: phonetic resistance to structurally-conditioned pitch accent reduction (Japanese).
Gesture coordination is sensitive to linguistic context: **collaborators**

Japanese

CVC $\rightarrow$ CC

Mandarin

CVer $\rightarrow$ CV

Tibetan

CVC $\rightarrow$ CC

Shigeto Kawahara
Keio University

Muye (Andy) Zhang
Yale, PhD Candidate

Chris Gesissler
Yale, PhD Candidate
High vowel devoicing in Tokyo Japanese

*high vowels are devoiced between two voiceless consonants and between a voiceless consonant and a pause*

| ōtaise: | ‘individuality’ |
| jisen | ‘eye gaze’ |
| ōsoku | ‘shortage’ |
| tʃikai | ‘pledge’ |
| katsu toki | ‘win time’ |
| aʃika | ‘sea lion’ |
| ōtaise: | ‘individuality’ |
| jisen | ‘eye gaze’ |
| ōsoku | ‘shortage’ |
| tʃikai | ‘pledge’ |
| katsu toki | ‘win time’ |
| aʃika | ‘sea lion’ |
| ōtaise: | ‘individuality’ |
| jisen | ‘eye gaze’ |
| ōsoku | ‘shortage’ |
| tʃikai | ‘pledge’ |
| katsu toki | ‘win time’ |
| aʃika | ‘sea lion’ |

The lingual gesture of devoiced vowels is optionally deleted

Largely categorical results—most tokens either [u] or [∅]

Bayesian classification of devoiced trajectories

CC vs. CVC

What happens to gestural coordination when the vowel height target for /u/ is deleted?

CC sequence
C₂ is timed to the release of C₁

C₂ \text{Tar} = C₁^{\text{Rel}} + k^{ipi} + \varepsilon

CVC sequence
C₂ is timed to the end of V

C₂ \text{Tar} = V^{\text{End}} + k^{elo} + \varepsilon

vowel

V^{\text{End}}
Simulation algorithm (effect of $C_1$ plateau duration on IPI)

**CC sequence**

$C_1^{Tar} = C_1^{Rel} - kp + \epsilon$

$C_1^{Rel} = N(\mu, \sigma^2)$

$C_2^{Tar} = C_1^{Rel} + k_{ipi} + \epsilon$

Varied from 75:125 ms in 1 ms steps

**CVC sequence**

$C_1^{Tar} = C_1^{Rel} + kp + \epsilon$

$C_1^{Rel} = N(\mu, \sigma^2)$

$C_2^{Tar} = V^{End} + k_{clo} + \epsilon$

$V^{Start} = C_1^{Tar} - k_{lag} + \epsilon$

$V^{End} = V^{Start} + k_{vdur} + \epsilon$
Simulation results

CC (vowel absent)  CVC (vowel present)

Difference in inter-consonantal interval

Duration of first consonant plateau (in ms)
Japanese data


6 speakers; 10-15 reps per item

[mastaː] [ɸsoku]  
[ʃtaiseː] [katstoki]  

[masuːtaː] [ɸusoku]  
[ʃuːtaiseː] [katsuːtoki]  

<table>
<thead>
<tr>
<th>Duration of the first consonantal gesture (in ms)</th>
<th>Inter-consonantal interval (in ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td><img src="image" alt="Graph showing the relationship between consonantal duration and inter-consonantal interval." /></td>
</tr>
<tr>
<td>absent</td>
<td><img src="image" alt="Graph showing the relationship between consonantal duration and inter-consonantal interval." /></td>
</tr>
</tbody>
</table>

C₁

C₂

vowel
Replication

6 new speakers; more items; 10-15 reps

Only coronal-initial cluster showed gestural reorganization.

Discussion: discontinuous variation

• In Tokyo Japanese, devoicing triggers variable deletion (categorical) of a vowel height target in [u].

• Deletion of vowel height target triggers gestural re-organization (categorical) for [ɸ]-initial words but not for [ʃ]-initial words.

• Possibly related to lexical gap:

<table>
<thead>
<tr>
<th></th>
<th>/u/</th>
<th>/i/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ꙑ</td>
<td>Ꙟu</td>
<td>ꙟi</td>
</tr>
</tbody>
</table>

• Gestural reorganization except when contrast is at stake?
Gesture complexity and tone

• So far, we’ve looked at rather simple distinctions:

• Combinations of coordination relations can apply competing forces, resolved by compromise.

• In this respect tone has been observed to behave similarly to segmental gestures.
Tone as gesture

Russian (Kozhevkikov & Chistovich, 1965)
American English (Lofqvist & Gracco, 1999)

Romanian (Marin, 2013),
American English (Marin & Pouplier, 2010),
Polish (Hermes et al., 2017)

Mandarin Chinese (Gao, 2008; Shaw & Chen, 2019)
Thai (Karlin & Tilsen, 2015)
Lhasa Tibetan (Hu, 2016)
Mandarin Chinese tones

**Lexical tones** (1-high, 2-rising, 3-low, 4-falling)


**Toneless syllables** or “neutral tone” (Chen & Xu, 2006)

- lexically toneless, e.g. *ma* QUEST, *le* PERF → “absent”
- disyllabic words, e.g. /bō.lí/ ‘glass’ > [bō.li], /yún.cǎi/ ‘cloud’ > [yún.cai] → “reduced”
  especially in compounds, e.g. [bō.li.bēi] glass cup
Mandarin materials – 7 sets

- Participants read context silently and then read aloud target words in sentences and in isolation.
Methods

• Electromagnetic Articulography (EMA)
  • upper/lower lip sensors => closure in /m/ gesture
  • tongue dorsum sensor => retraction in /u/ gesture

• 11 participants
  • 6 female
  • ages 19-37 (mean 22;4)
  • native speakers of mandarin

• 6,798 tokens
  (2 pronunciations (sentence/isolation) x 3 conditions
  (full/reduced/absent) x 7 sets = 42 tokens per block; 12-19
  blocks per participant)

• Tone presence/absence determined by Bayesian classification (Shaw & Kawahara 2018).
Results

Tone presence/absence based on f0

Results

full-tone

absent-tone (phonetic interpolation)

Tone presence/absence based on f0

 condition

reduced
absent
full

condition

reduced
absent
full

full-tone

absent-tone

m

u

posterior probability

short lag
n.S.

long lag

z-scored C-V lag (% of syllable duration)

***
Discussion

• As expected, long CV lag for full tone syllables; short lag for neutral tone

• Surprisingly, “Reduced” syllables showed full tone pitch trajectory but short CV lag

• Morpho-syntactic context triggers shift in gestures; tone undershoot/loss follows
diaspora Tibetan

• Derives from a mix of Tibetan varieties some of which have lexical tone and some of which do not.

• Tonal dialects have two-way tonal contrast:
  • High tone (H)
  • Rising tone (LH)
Methods

• Electromagnetic Articulography (EMA)
  • upper/lower lip sensors => closure in \[ p \ p^h \ m \]
  • tongue dorsum sensor => retraction in \[ u \ o \ a \]

• 6 participants
  • 4 female
  • ages 19-37
  • native speakers of diaspora Tibetan

• 3,862 tokens for analysis
  • 72 items read in carrier phrase:
  • 5-10 reps per item

\[ \text{dependence measure} \]

\[ \text{onset-to-onset} \]

\[ \text{CV lag} = \text{onset-onset} \]

‘This word is ___.’
Results

• 4 speakers produce tone contrast; 2 do not
• All 6 speakers show long CV lag
Discussion

• “Lexical tone languages” tend to have long lag C-V timing

• Even when they’ve lost tone (Tibetan) the timing pattern can persist in the community, indicating that it is not the presence of the tone per se that conditions long lag (synchronously)

• Likewise in Mandarin it is not the loss of tone that triggers synchronous timing (in “reduced” condition), but rather the synchronous timing that causes tone undershoot (and ultimate loss)
Devoicing triggers deletion of vowel height target which triggers change in coordination.

Context (morpho-syntactic or maybe prosodic) triggers change in gesture coordination which leads to tone undershoot.
This talk

1) **Dynamic invariance**: still a good idea!
   – Gestural basis for complex segments (Russian, English)

2) Gestural coordination is **conditioned by linguistic context**
   – Gesture deletion triggers re-organization of gestural coordination (Japanese)
   – Re-organization of gestural coordination precipitates tone loss (Mandarin)
   – Tone loss proceeds without gestural re-organization (diaspora Tibetan)

3) **Living lexicon: word-specific phonetics**
   – **Lexical absorption**: words take on the phonetic detail of the prosodic environments in which they are typically produced (Mandarin)
   – **Lexical persistance**: phonetic resistance to structurally-conditioned pitch accent reduction (Japanese).
Living lexicon

• **Lexical persistence**: resistance to structurally conditioned reduction (Kawahara, Shaw, Ishihara, 2021)

• **Lexical absorption**: lexical items take on the phonetic detail of the prosodic environments in which they are typically produced (Tang & Shaw, 2021)
Lexical absorption: Prosody leaks in the lexicon


Kevin Tang  
U. of Florida
The argument

- **Prosodic prominence**, as dictated by contextual predictability, influences *word duration, pitch, and intensity* in Mandarin Chinese.

- A word's **informativity** (average contextual predictability) reflects aggregate influences of prosodic prominence on lexical representations.

- **Informative words** have longer duration, higher pitch, greater intensity, *even in prosodically weak positions*, because they tend to occur in prosodically prominent positions.

- Hence, prosodic prominence leaks into the lexicon.
Living lexicon: feedback loop
Lexical persistence: Failure of prosodic reduction

A puzzle in syntactic theory

• Some languages (e.g. Tagalog) show **overt wh-movement**; some languages allow their **wh-elements to stay in situ** (e.g. Japanese).

• **Minimalist Syntax**: those that move overtly have a strong (uninterpretable) feature that needs to be checked. Japanese wh-elements on the other hand have a weak (interpretable) feature.

• **Richards (2010)** attempts to derive this difference from an independently observable difference. (Further developed in Richards 2016).

Richards’ (2010) proposal in a nutshell

• All languages attempt: “to create a prosodic structure for wh-questions in which the wh-phrase and corresponding complementizer are separated by as few prosodic boundaries as possible” (p. 145).

• Japanese has a prosodic means to group the wh-phrase and its complementizer, and hence does not need to resort to overt wh-movement.

• Tagalog on the other hand does not have that prosodic strategy, so its wh-elements needs to move overtly.
One source of inspiration for Richards (2010)

• Post-wh accent in Japanese is **eradicated** (Deguchi & Kitagawa 2002). Sample pitch tracks from Ishihara (2001).

(28a): Non-interrogative sentence

(28b): Wh-question

(29b): Wh-question

Figure 3-1: Single wh-question
Deguchi & Kitagawa (p.74)

“Another important prosodic effect of focus pointed out by Ishihara (2000) (extending the original observation by Ladd (1996)) is that an emphatic accent is accompanied by what we label as "eradication" of lexical accents. That is, when one or more of lexical accents follow an emphatic accent, their H tones (H*) are all suppressed. As a result, the lowest pitch induced by the emphatic accent is inherited and prolonged with further gradual declination up to the right boundary of some clausal structure”
Method

- **Tone presence/absence** determined by **Bayesian classification** (Shaw & Kawahara 2018).
- Nine Tokyo Japanese speakers (4 female)
- 6 items per condition; 2 repetitions each (24 tokens per subject)

(1) Control sentences: Word₁ Word₂[-wh] Word₃ Word₄ Verb
(2) Test sentences: Word₁ Word₂[+wh] Word₃ Word₄ Verb

(1) 丸山はエルメスの襟巻きに飲み物をこぼしました。Maruyama-TOP Hermes-GEN scarf-DAT drink-ACC spilled
(2) 丸山はどの人の襟巻きに飲み物をこぼしましたか？Maruyama-TOP Who scarf-DAT drink-ACC spilled-Q

Clear tokens of tone eradication

Full tone (same as [-wh])

Reduced tone

All speakers

Posterior probability of targetlessness
Results by speaker

Every speaker produces some tokens without reduction
Living lexicon: direct actuation

Japanese
Prosodic context conditions tone eradication, but lexically specified pitch accents can still get through.
General discussion

- Different gestural coordination patterns can be distinguished in the kinematics because they structure variation in specific ways (dynamic invariance), but there’s more...

- **Context**, including prosodic context, conditions gestural reorganization and can **feedback** into the lexicon.

- Contextual factors can also be bypassed, c.f. motor program reuse.
Thank you!
EXTRA SLIDES
Approach

• Setup stochastic generators of $f_0$ based on competing phonological hypotheses:
  \[ H_1: \text{LHL} \]
  \[ H_2: \text{LØL} \]

• Use stochastic generative model to assign probabilities of phonological hypotheses to phonetic data.

• Allows for token-by-token analysis of $f_0$ contours

Which phonological structure is responsible for the phonetic data?
**Step 1: Discrete Cosine Transform (DCT)**

Represent \( f_0 \) trajectory as the sum of Cosines:

\[
y(k) = w(k) \sum_{n=1}^{L} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)
\]

\( k = 1, 2, \ldots, L \)

Where \( L \) is the number of data samples and \( x(n) \) is the trajectory to be modelled and:

\[
w(k) = \begin{cases} 
\frac{1}{\sqrt{L}} & k = 1 \\
\sqrt{\frac{2}{L}} & 2 \leq k \leq L 
\end{cases}
\]

Fit between real and simulated F0 using iDCT

- Simulations from 4 DCT components explain > 90% of variance for all 9 speakers
Step 2: F0 of LØL (the noisy null)

Simulate F0 trajectories from DCT components:

Interpolation trajectory \( y(k) \sim N(\mu(k), \sigma(k)) \)

Target present [-Wh] \( y(k) \sim N(\mu(k), \sigma(k)) \)

\[
x(n) = \sum_{n=1}^{L} w(k)y(k) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)
\]

\( n = 1, 2, \ldots, L \)

Where \( L \) is the number of data samples and \( x(n) \) the trajectory to be simulated and:

\[
w(k) = \begin{cases} 
\frac{1}{\sqrt{L}} & k = 1 \\
\sqrt{\frac{2}{L}} & 2 \leq k \leq L 
\end{cases}
\]
Step 3: Bayesian classifier

- Training data
  - [-Wh] Word3 Word4
  - Linear interpolation

- Test data
  - [+Wh] Word3 and Word4

\[
p(T \mid C_{o_1}, \ldots, C_{o_n}) = \frac{p(T) \times \prod_{i=1}^{n} p(C_{o_i} \mid T)}{\prod_{i=1}^{n} p(C_{o_i})}
\]

where \( C_{o_i} \) is the \( i \)th DCT coefficient
English: By subject

- **F03**
  - Equation: $y = -107 + 0.642x$
  - $R^2 = 0.27$

- **F04**
  - Equation: $y = -72.9 + 0.425x$
  - $R^2 = 0.17$

- **M01**
  - Equation: $y = -75.6 + 0.513x$
  - $R^2 = 0.26$

- **M02**
  - Equation: $y = -114 + 0.707x$
  - $R^2 = 0.26$

**Axes:**
- **Y-axis:** Difference in gestural onset time (ms)
- **X-axis:** Duration of the first consonant gesture (ms)
# Methods: Russian

## Speakers

Four native speakers of Russian (3 female and 1 male)

## Stimuli

<table>
<thead>
<tr>
<th>Target</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ci</td>
<td>Cj</td>
</tr>
<tr>
<td>[pʲok]</td>
<td>[pjot]</td>
</tr>
<tr>
<td>[bʲust]</td>
<td>[bjut]</td>
</tr>
<tr>
<td>[mʲu]</td>
<td>[mju]</td>
</tr>
<tr>
<td>[fʲodor]</td>
<td>[fjord]</td>
</tr>
<tr>
<td>[vʲoz]</td>
<td>[vjoš]</td>
</tr>
<tr>
<td>[vʲodra]</td>
<td>[vjotsa]</td>
</tr>
</tbody>
</table>

Carrier phrase: [ʌˈna ___ pəftʲɪlʲilʲ]. ‘She ____ repeated.’
# Methods: English

## Speakers
Four native speakers of English (2 female and 2 male)

## Stimuli

<table>
<thead>
<tr>
<th>Target</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td>butte</td>
<td>frap</td>
</tr>
<tr>
<td>pew</td>
<td></td>
</tr>
<tr>
<td>view</td>
<td></td>
</tr>
<tr>
<td>mew</td>
<td></td>
</tr>
<tr>
<td>musical</td>
<td></td>
</tr>
</tbody>
</table>

- Carrier phrase: ‘It’s a ____ perhaps.’