Consistent C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts

Christopher Geissler\textsuperscript{1}, Jason Shaw\textsuperscript{1}, Mark Tiede\textsuperscript{2}, Fang Hu\textsuperscript{3}

\textsuperscript{1}Yale University, \textsuperscript{2}Haskins Laboratories, \textsuperscript{3}Chinese Academy of Social Sciences, Beijing

This material is based upon work supported by the National Science Foundation under Grant Number 1928750
Introduction: Gestures

- Articulatory gestures: abstract, dynamic representations of controlled movements of the vocal tract
  
  (e.g. Browman & Goldstein 1986)

- How are gestures timed with each other?

“C-V lag”

<table>
<thead>
<tr>
<th>Tongue</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>front</td>
<td>↓</td>
<td>back</td>
</tr>
<tr>
<td>Dorsum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>open</td>
<td>↓</td>
<td>closed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lips</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questions

• How are speech gestures timed with each other?

  • Which gestures are encoded in the phonological representation?

• How are those gestures coordinated?
Which gestures?

- Strong Articulatory Phonology hypothesis: only gestures encoding phonological contrasts are represented

- C, V: specified for constriction location and degree
  
  (e.g. Browman & Goldstein 1986)

- tones: relative F0 excursions
  
  (Gao 2008, Katsika et al 2014)
How are gestures coordinated?

- *Planning oscillators* coordinate gestures with cyclic phasing.

- In-phase (0°) and anti-phase (180°) coupling modes learned more easily, as in general motor coordination (*Browman & Goldstein 2000, Saltzman et al 2008)*.

- Other phasing modes are possible, but more difficult (eccentric timing) (*Goldstein 2011)*.
Two stable coupling modes

- in-phase → synchronous start times
- anti-phase → sequential start times
Exceptional C-V timing

- Onset clusters often show partial overlap ("C-center")
  (Browman & Goldstein 1988, inter alia)
- Exceptional clusters:
  - some CV timing unchanged when add earlier C:
    - Italian /sC/ onsets (Hermes et al. 2008, 2011)
    - Moroccan Arabic (Shaw et al 2009), Tashlhyit Berber
      (Goldstein et al. 2007, Hermes et al. 2017)
Explanation: competitive coupling

- Account for partial overlap through competition between in-phase and anti-phase coupling:
  
  \( (\text{Browman & Goldstein 2000, Nam & Saltzman 2003}) \)
C-V timing with tone

• Intonational tone doesn’t affect C-V timing in some languages: German and Italian (Niemann et al. 2011), Catalan (Mücke et al. 2012). C-V lag <10ms

• Longer C-V lag in Mandarin (Gao 2008), Thai (Karlin 2014), and Lhasa Tibetan (Hu 2016). C-V lag ~50ms

• Toneless syllables in Mandarin show reduced C-V lag relative to their fully-tonal counterparts (Zhang et al. 2019)
C-V lag and tone

- Different potential structures for tone gestures:
  - Mandarin
  - Thai
  - Tibetan

  [Diagram]

  [Diagram]

  Italian
  German
  Catalan
Evidence for tone ~ C-V lag

• Tone is associated with longer C-V lag:
  • in lexical tone languages  
    vs. non-lexical tone languages
  • in tonal and toneless syllables  
    in the same language
  • present study: across speakers with  
    vs. without tone contrast in the same language
Hypothesis

- In a language where some speakers produce a tone contrast and others do not:

  tone-contrasting speakers $\rightarrow$ positive C-V lag
  non-tone-contrasting speakers $\rightarrow$ near-zero C-V lag
Tibetan

- Tonal and non-tonal dialects

- Tone contrast: H vs. LH
  
  *(Duanmu 1992, Tournadre and Dorje 2003)*

- Speakers raised in post-1959 diaspora (India, Nepal) exposed to mixed input, acquire mixed features
  
  *(Geissler 2018)*
EMA Experiment

- Electromagnetic Articulography (EMA) to track fleshpoints on the lips and tongue, alongside audio
  - goal: quantify timing of oral gestures
- 6 Tibetan speakers (4 female) raised in Diaspora
  - all multilingual, extensive dialect contact
Methods

• Speakers read words in carrier phrase on a screen, in Tibetan orthography

• EMA sensors on each lip and three on tongue; head movement corrected w/r/t/three sensors on rigid points of the head

• Gesture start labelled at 20% of peak velocity to target
Stimuli

- Bilabial onsets: separate C and V articulators
- Back vowels following front vowel in consistent frame sentence
- \(/m \, p \, p^h/ \) * 2 tones * /a \, o \, u/ * CV/CVC syllables * mono/disyllabic * 10 repetitions
Identifying tone contrast

• measured F0 at ten time-normalized points along [mV] syllables (60 words per speaker)

• acoustic analysis in Praat *(Boersma and Weenink 2018)*; VOT and time-normalized pitch calculated using Praat scripts *(DiCanio 2011, 2018)*
Results: tone contrast

- fitted GAMMs to predict F0 based on:
  - parametric term for tone
  - smooth for timestep at reference value for tone
  - difference smooth across tones
  - random smooths by word
Results: tone contrast

- 4 speakers produce a tone contrast, two do not (/mV/)
Results: tone contrast

- Confirm with GAMM (smooths for tone plotted)
## Results: tone contrast

<table>
<thead>
<tr>
<th>term</th>
<th>F01</th>
<th>F02</th>
<th>F03</th>
<th>M02</th>
<th>M01</th>
<th>F04</th>
</tr>
</thead>
<tbody>
<tr>
<td>tone (parametric)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time smooth</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>difference smooth by tone</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>random smooths by word</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Deviance explained</td>
<td>67.8%</td>
<td>94.8%</td>
<td>80%</td>
<td>71.6%</td>
<td>77%</td>
<td>8.57%</td>
</tr>
</tbody>
</table>
Results: C-V lag

- Among tone-contrasting speakers, C-V lag is positive!
Results: C-V lag

• ... but also for non-contrasting speakers
Results: C-V phasing

- C-V lag relative to C duration also similar for speakers with and without tone contrast
Results: C-V phasing

- Confirmed results with comparison of LMMs:
  - baseline model: fixed effect of onset, random effects of speaker and word
  - comparison: baseline plus fixed effect of tone contrast

<table>
<thead>
<tr>
<th>model</th>
<th>Df</th>
<th>AIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>6</td>
<td>12461</td>
<td>-6224.5</td>
</tr>
<tr>
<td>comparison</td>
<td>7</td>
<td>12462</td>
<td>-6223.7</td>
</tr>
<tr>
<td>Tone contrast</td>
<td>Predicted</td>
<td>Observed</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="predicted_tone_contrast" alt="Diagram" /></td>
<td><img src="observed_tone_contrast" alt="Diagram" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No tone contrast</th>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="predicted_no_tone_contrast" alt="Diagram" /></td>
<td><img src="observed_no_tone_contrast" alt="Diagram" /></td>
</tr>
</tbody>
</table>

26
<table>
<thead>
<tr>
<th>Tone contrast</th>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="predicted_tone_contrast.png" alt="Diagram" /></td>
<td><img src="observed_tone_contrast.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No tone contrast</th>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="predicted_no_tone_contrast.png" alt="Diagram" /></td>
<td><img src="observed_no_tone_contrast.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Summary

• Some diaspora speakers contrast tone, others don’t

• Observe long C-V lag in Tibetan, like Thai and Mandarin

• ... but speakers with and without tone show similar C-V lag

• C-center-like timing learned even by speakers lackng the tone contrast
Interpretation

• How to account for similar C-V lag across speakers with and without tone production contrast?

• Possibility 1: Non-contrasting speakers have a non-contrastive (tone?) gesture

• Possibility 2: Non-contrasting speakers use eccentric C-V timing
Interpretation

• How account for similar C-V lag across speakers with and without tone production contrast?

• Possibility 1: Non-contrastting speakers have a non-contrastive (tone?) gesture

• different from “Strong AP hypothesis” where only contrastive gestures are in the coupling graph
Interpretation

• How account for similar C-V lag across speakers with and without tone production contrast?

• Possibility 2: Non-contrasting speakers learn the same C-V timing spoken around them

• without competitive coupling, as eccentric timing

(e.g. Marin & Pouplier 2010, Goldstein 2011)
Conclusions

• Tibetan speakers with and without a tone production contrast showed similar C-V lag

• Speakers can learn eccentric timing relations resembling those of other members of the speech community

• Eccentric timing can resemble competitive coupling
References I


References II


C-V lag by tone

- No effect of tone on C-V lag

Density plot of C-V lag

Density plot of C-V phasing
Results: C-V timing

- C-V lag not significantly different by aspiration either (/pV/ vs. /pʰV/)

- LMM: random effects of speaker, word; fixed effect of tone contrast

- model not improved by adding effect of onset
Results: C-V timing

- C-V lag not significantly different across tones (/mV/)

![C-V lag by speaker and tone graph](image)
Results: C-V Phasing / Cdur

Effect of C duration on C–V lag (normalized)
Tibetan

- Tonal and non-tonal dialects

- Tonal dialects:
  - One tone per word
  - Two tone type: high-level and low-rising

\[
\begin{align*}
\sigma & \quad \sigma_1 & \quad \sigma_2 \\
H & \quad H \quad H & \quad L \quad H \quad L \quad H
\end{align*}
\]