Synthesis from constriction tasks

- Different consonants and vowels can be produced (by humans and by a synthesizer) with different combinations of a small number of simple constriction tasks in the vocal tract.

- Each task is performed by a dynamical system, whose goal and stiffness is specific to that task, via a synergy of articulators.

- The tasks are arranged in time.

- All of the words we know, and ones we have never said before can be produced by combining tasks over time.

- How do we know what tasks to use for a novel word?
  - “plitch”

- TaDA can produce any word in a similar way.

- For now, in this class, we have to tell TaDA explicitly what the tasks are for some word, but there is also a model that know what list of tasks (gestural score) to use for a word (or novel word).
Synthesis in different languages

- All languages can be synthesized in the same way, with similar tasks.
- The exact task goals may differ from language to language.
- There are a few tasks that are employed in other languages, but not in English.
- The way tasks are arranged in time can be quite different.
TADA Synthesizer

- TaDA synthesizes speech from a list of constriction tasks (also called *gestures*) to be produced by dynamical systems.
- Each task is specified for:
  - Identity of Task variable (state)
  - Interval of time during which the dynamics of this task controls the task variable
  - Goal value
  - Stiffness
  - Articulators that form the synergy for that task, and their average relative contributions or weights (big numbers mean the articulator is “heavy” and contributes less).
- Blending parameters ($\alpha$ and $\beta$ for when more than one Task controls the same variable at the same time (overlap, like we saw with the thermostat with leaky insulation)).

<table>
<thead>
<tr>
<th>Task</th>
<th>Beg Fr</th>
<th>End Fr</th>
<th>Goal</th>
<th>k</th>
<th>Articulator Weights</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>'LA'</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>-2</td>
<td>8 1  JA=8, UH=5, LH=1</td>
<td>100</td>
<td>0.01</td>
</tr>
</tbody>
</table>
The articulators available to the Tasks in TADA are the CASY articulators.

TADA incorporates knowledge of the relation between articulator positions and the values of the task variables.

As the task variable changes over time (due to the dynamical system), a model of the synergy calculates the change in CASY articulator positions.

Given articulator positions, CASY calculates vocal tract shapes, areas and formants.
### TADA Synthesizer: Articulator Weights

<table>
<thead>
<tr>
<th>Task</th>
<th>Beg Fr</th>
<th>End Fr</th>
<th>Goal</th>
<th>k</th>
<th>Articulator Weights</th>
<th>α</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>'LA'</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>-2</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

- LH (Lower Lip Height) will move the most.
- UH (Upper Lip Height) will move the next most.
- JA (Jaw Angle) will move the least.

Everything else being equal:
Full set of gestures in the production of /b/

%/b%/  
' LA' 0 0 9 0 -2 8 1 JA=8,UH=5,LH=1 100 0.01  
' LA' 0 9 13 0 11 8 1 JA=8,UH=5,LH=1 1 1  
' VEL' 0 0 9 0 -0.1 8 1 NA=1 0 0

- LIP CLOSURE
- LIP RELEASE
- VELIC CLOSURE
Synthesizing an entire word

- We need to know what the tasks are for vowels and other consonants.
- We need to know how to arrange them in time.
- Let’s try to synthesize the word “bob”.
- Need find the task for vowels: Constriction Location and Degree of tongue body.
TBCL and TBCD for Vowels

TBCL = 95
TBCD = 4

TBCL = 170
TBCD = 11

IY

AA
TADA Synthesizer: CASY Tongue Articulators for TB tasks
Full set of gestures in the production of /aa/

%/aa/
'TBCD' 0 0 30 0 11 4 1 JA=1,CL=1,CA=1 1 1 1
'TBCL' 0 0 30 0 170 4 1 JA=1,CL=1,CA=1 1 1 1
Putting tasks together in time

- The tasks for the initial C and the V begin at the same time!
- The V task continues about twice as long as the C task
- The task for the final C begins after the V ends.
Time to get to goal in x-ray data: C vs. V

“two back”

- C gets to its goal faster than V:
- $k=8$ for consonant gestures
- $k=4$ or consonant gestures
Full set of gestures for /b/, /aa/ 

%/b%/  
'LA' 0 0 9 0 -2 8 1 JA=8, UH=5, LH=1 100 0.01  
'LA' 0 9 13 0 11 8 1 JA=8, UH=5, LH=1 1 1  
'VEL' 0 0 9 0 -0.1 8 1 NA=1 0 0  

%/aa%/  
'TBCD' 0 0 30 0 11 4 1 JA=1, CL=1, CA=1 1 1  
'TBCD' 0 0 30 0 170 4 1 JA=1, CL=1, CA=1 1 1 1
TADA input/output for “bob”: Gestural Score

```
% /b/
'LA' 0 0 9 0 -2 8 1 JA=8, UH=5, LH=1 100 0.01
'LA' 0 9 13 0 11 8 1 JA=8, UH=5, LH=1 1 1
'VEL' 0 0 9 0 -0.1 8 1 NA=1 0 0

% /b/
'TBCD' 0 0 30 0 11 4 1 JA=1, CL=1, CA=1 1 1
'TBCD' 0 0 30 0 170 4 1 JA=1, CL=1, CA=1 1 1

% /a/
'TBCD' 0 0 30 0 11 4 1 JA=1, CL=1, CA=1 1 1
'TBCD' 0 0 30 0 170 4 1 JA=1, CL=1, CA=1 1 1
```

```
0 50 100 150 200 250 300 350
G(LA) = -2  G(LA) = -2  G(LA) = 11  G(LA) = 11
G(VEL) = -0.1  G(VEL) = -0.1
G(TBCD) = 11
```

October 19, 2015
Change “bob” to “mob”?

- Change initial VEL goal from -0.1 to 0.2

```
% /m/
'LA' 0 0 9 0 -2 8 1 JA=8, UH=5, LH=1 100 0.01
'LA' 0 9 13 0 11 8 1 JA=8, UH=5, LH=1 1 1
'VEL' 0 0 9 0 0.2 8 1 NA=1 0 0
%
% /aa/
'TBCD' 0 0 30 0 11 4 1 JA=1, CL=1, CA=1 1 1
'TBCL' 0 0 30 0 170 4 1 JA=1, CL=1, CA=1 1 1
%
% /b/
'LA' 0 25 34 0 -2 8 1 JA=8, UH=5, LH=1 100 0.01
'LA' 0 34 37 0 11 8 1 JA=8, UH=5, LH=1 1 1
'VEL' 0 25 34 0 -0.1 8 1 NA=1 0 0
```
"bob" vs. "mob"

- All the tasks remain the same, but one!
- We can combine tasks into new combinations the way we cannot with pieces of sound ("stick")
How do we change “bob” to “pop”?

Goal for $GLO = 0.4$ for voiceless stops and fricatives

- Only one articulator for this task ($GW$)
- Neutral $= 0.0$
All the tasks remain the same, we just add a new one.
### Other Tasks

<table>
<thead>
<tr>
<th>Segments</th>
<th>Task Variables</th>
<th>Articulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>p b m</td>
<td>Lip Aperture (LA)</td>
<td>Upper Lip (uh) Lower Lip (lh) Jaw (ja)</td>
</tr>
<tr>
<td>t d n s z sh zh</td>
<td>Tongue Tip: Constriction Degree (TTCD) Constriction Location (TTCL)</td>
<td>Tongue Tip (ta, tl) Tongue Body (cl, ca) Jaw (ja)</td>
</tr>
<tr>
<td>k g nx Vowels</td>
<td>Tongue Body: Constriction Degree (TBCD) Constriction Location (TBCL)</td>
<td>Tongue Body (cl, ca) Jaw (ja)</td>
</tr>
<tr>
<td>m n nx</td>
<td>Velic Aperture (VEL)</td>
<td>Velum (na)</td>
</tr>
<tr>
<td>p t k s sh</td>
<td>Glottal Aperture (GLO)</td>
<td>Glottal Width (gw)</td>
</tr>
<tr>
<td>uw uh ow ao</td>
<td>Protrusion (PRO)</td>
<td>Lip Protrusion (lp)</td>
</tr>
</tbody>
</table>

---

**Diagram:**

- **Segments:** p b m, t d n s z sh zh, k g nx, m n nx, p t k s sh, uw uh ow ao
- **Task Variables:** Lip Aperture (LA), Tongue Tip, Tongue Body, Velic Aperture, Glottal Aperture, Protrusion
- **Articulators:** Upper Lip (uh), Lower Lip (lh), Jaw (ja), Tongue Tip (ta, tl), Tongue Body (cl, ca), Velum (na), Glottal Width (gw), Lip Protrusion (lp)
Tongue Tip Tasks

- For consonants: t d n

TTCD=-2
TTCL=56
TTCD=11
TTCL=24

release
closure
Articulators for TT Tasks
Articulators for TT Tasks
"beat"

Linguistics 285 (USC Linguistics)  Lecture 14: Speech Synthesis using TaDA 2   October 19, 2015  25 / 1
“bob” to “beat”

- Change $G(TBCD)$ to 11
- Change $G(TBCL)$ to 95
- Replace LA task with TTCD and TTCL tasks.
- Add GLO task
“map” vs. “nap”

- Replace LA task with TTCD and TTCL tasks.
TADA input/output for “ban” : Gestural Score

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Time (ms)</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>0 9 0 -2 8 1</td>
<td>JA=8, UH=5, LH=1, 100 0.01</td>
</tr>
<tr>
<td>LA</td>
<td>0 9 13 0 11 8 1</td>
<td>JA=8, UH=5, LH=1 1 1</td>
</tr>
<tr>
<td>VEL</td>
<td>0 9 0 -0.1 8 1</td>
<td>NA=1 0 0</td>
</tr>
<tr>
<td>/b/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ae/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/n/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

G(LA) = -2 G(LA) = 11
G(TTCD) = -2
G(TTCD) = 11
G(TBCD) = 17
G(VEL) = -0.1
G(VEL) = 0.2
“bin” vs. “ban”

`%/b/`
- 'LA' 0 0 9 0 -2 8 1 JA=8, UH=5, LH=1 100 0.01
- 'LA' 0 9 13 0 11 8 1 JA=8, UH=5, LH=1 1 1
- 'VEL' 0 0 9 0 -0.1 8 1 NA=1 0 0

`%/ih/`
- 'TBCL' 0 0 30 0 95 4 1 JA=1, CL=1, CA=1 1 1
- 'TBCD' 0 0 30 0 8 4 1 JA=1, CL=1, CA=1 1 1

`%/n/`
- 'TTCD' 0 34 37 0 11 8 1 JA=512, CL=512, CA=512, TL=1, TA=1 1 1
- 'TTCL' 0 34 37 0 24 8 1 JA=512, CL=512, CA=512, TL=1, TA=1 1 1
- 'VEL' 0 19 32 0 0.2 8 1 NA=1 1 1

- 'TTCD' 0 25 34 0 -2 8 1 JA=32, CL=32, CA=32, TL=1, TA=1 100 0.01
- 'TTCL' 0 25 34 0 56 8 1 JA=32, CL=32, CA=32, TL=1, TA=1 1 1
The line specifying lip closure task is the same for both [b]s.

- ’LA’ 0 0 9 0 -2 8 1  JA = 8, UH = 5, LH = 1  100  0.01
- ’LA’ 0 34 43 0 -2 8 1  JA = 8, UH = 5, LH = 1  100  0.01

But due to the low jaw in “ban”, the upper lip lowers more than it does in ”bin”, and the lower lip raises more.

This emerges automatically from synergy model.
TADA Synthesizer: “bin ban”

“bin”

“ban”

Lip Aperture
Upper Teeth to Upper Lip
Lower Teeth to Lower Lip
Jaw Angle

Time (ms)
Velar stop tasks