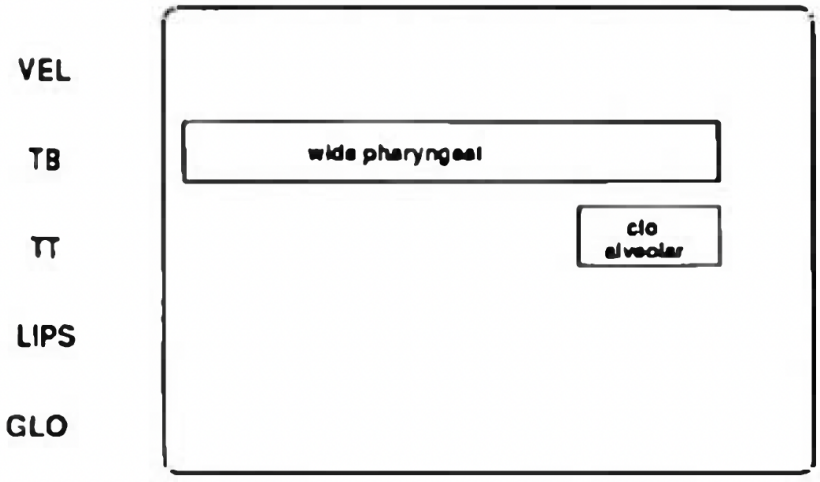
Lecture 10: Articulatory Phonology

Mark Hasegawa-Johnson

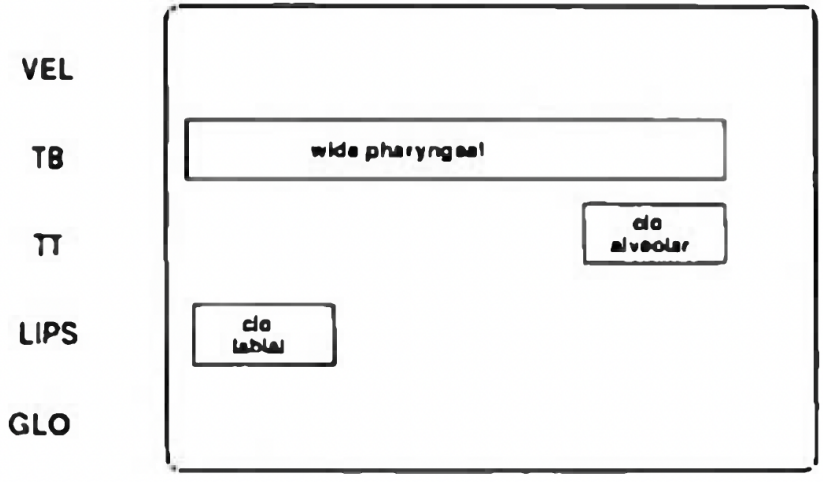
ECE 537: Speech Processing

1. Gestures instead of phonemes as the units of speech planning
2. Tract variables are learned abstract groupings of muscles
3. The discrete→continuous mapping is planned by a task dynamic model
4. Example
5. Summary

Outline



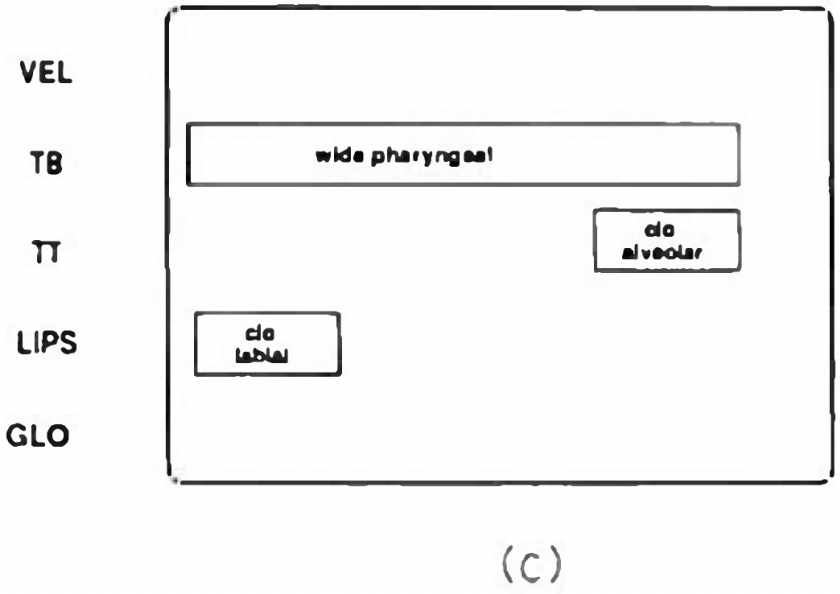
(a)



(c)

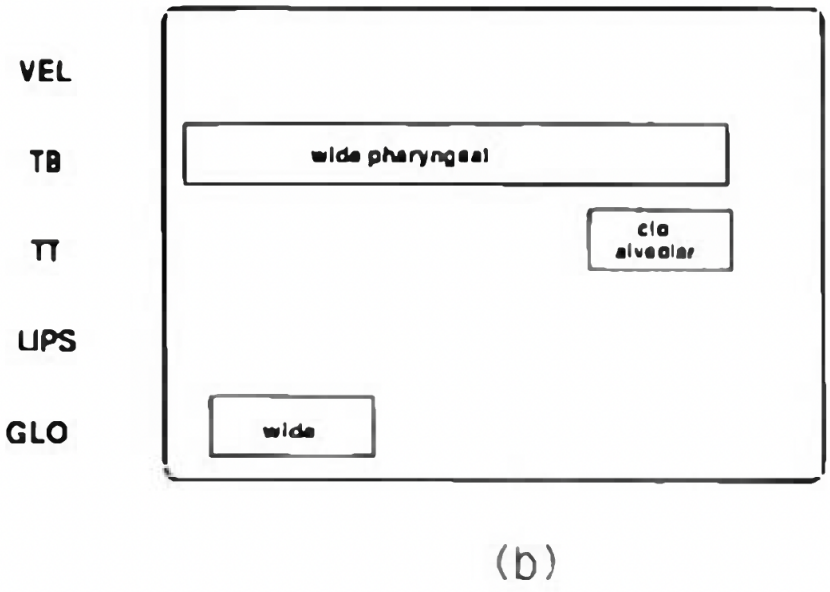
Gestures instead of phonemes

- The mental processes that generate speech start not with a sequence of phonemes, but with a constellation of gestures.
- Gestures change the default setting, e.g., the glottal default is "voiced."
- Difference between "add" and "bad," as shown, is absence/presence of the lip closing gesture.

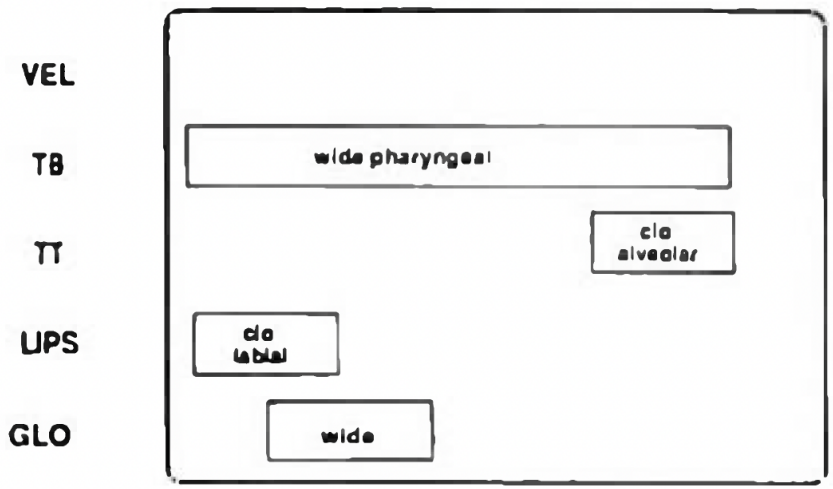


Primary articulators/Tract variables

- There are five primary articulators (called "tract variables"): Velum, Tongue Body, Tongue Tip, Lips, Glottis
- A closure gesture implemented by different articulators (lips vs. tip) results in different words (bad vs. dad)
- Each tract variable can only implement one gesture at a time



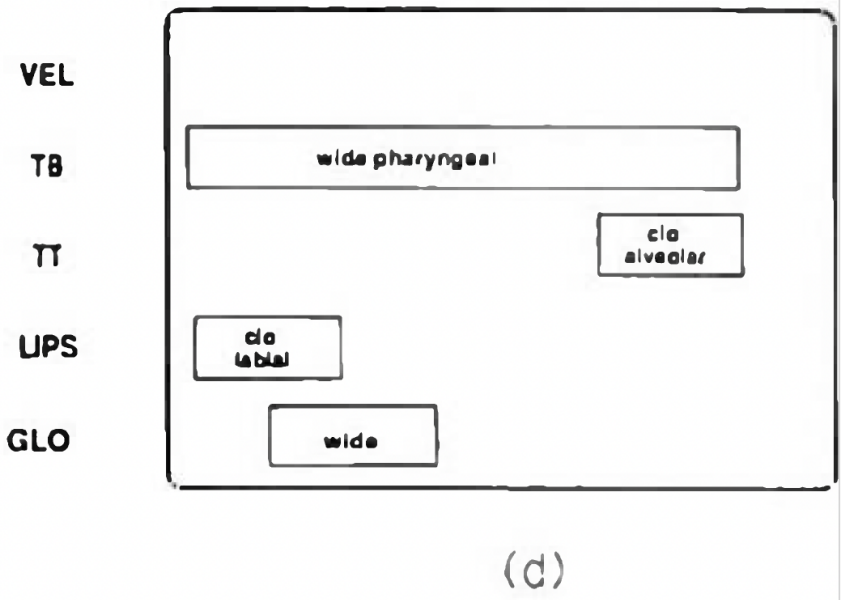
(b)



(d)

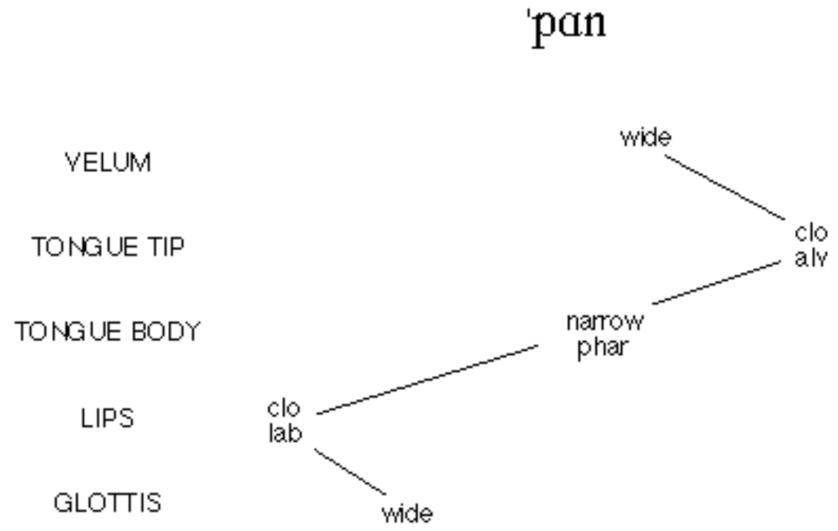
Synchronized gestures can change the phoneme

- A glottal opening gesture alone = [h]
- Glottal opening synchronized with the end of lip closure = [p^h]



Only one of each gesture at each time

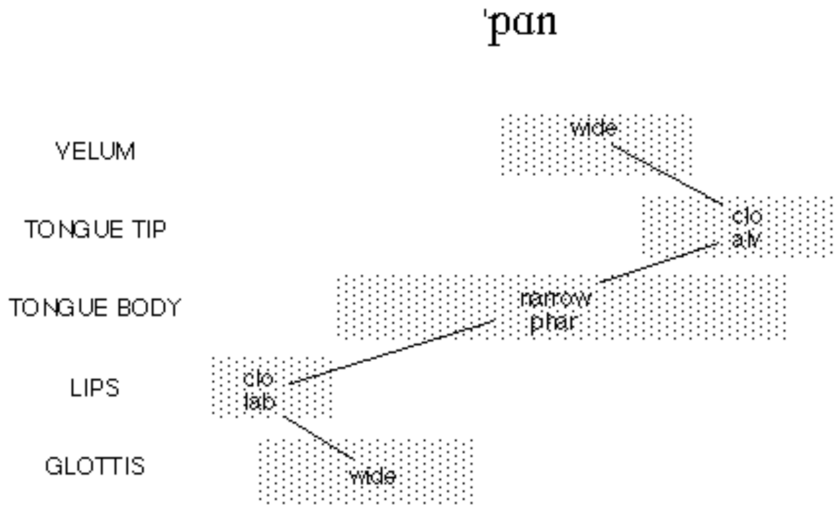
- Each gesture can occur only once in each syllable position
- Thus in the consonant cluster [sp] in "span," the [s] and [p] share the same glottal opening gesture



https://www.haskinslaboratories.org/gestural

Representations in memory are sparse

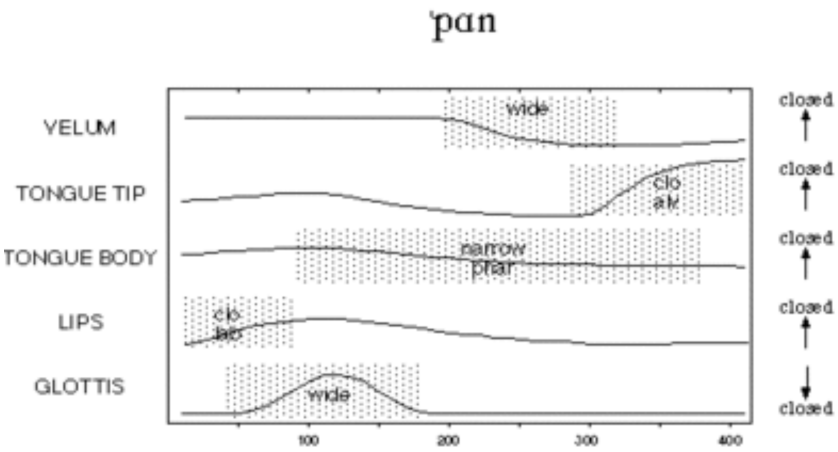
- A word in memory = synchronizations
- VEL WIDE $\leq$ TT C A
- L C L $<$ TB N P $<$ TT C A
- L C L $\leq$ G W



https://www.haskinslaboratories.org/gestural

Timing is part of planning

- Timing of gestures is computed once the word has been placed into a sentence
- Timing takes into account speaking rate and prosody



https://www.haskinslaboratories.org/gestural

Articulation is just movement

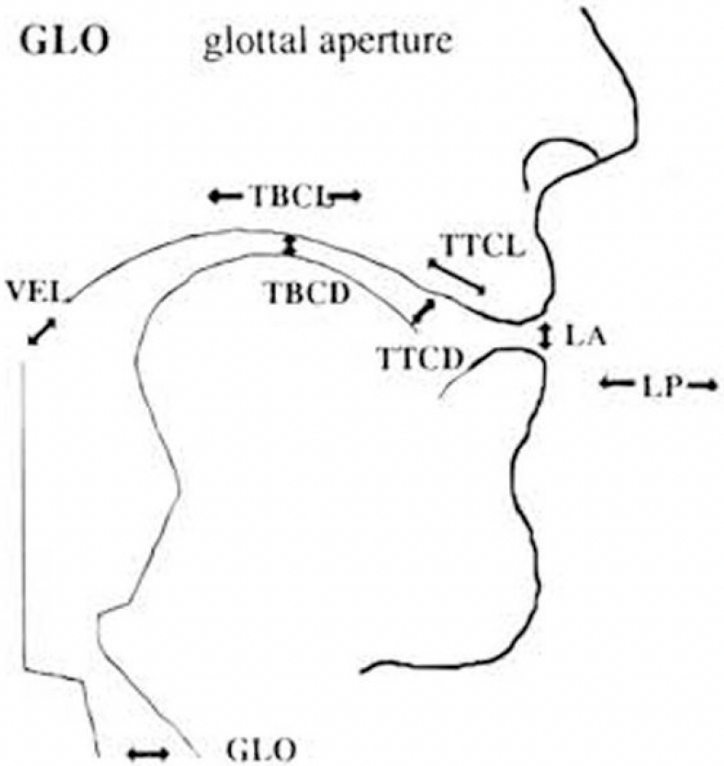
- Articulation is planned in the same way as any other movement
- Cognitive-internal world model plans movements to be efficient and smooth
- Once movements are planned, then commands are issued to muscles

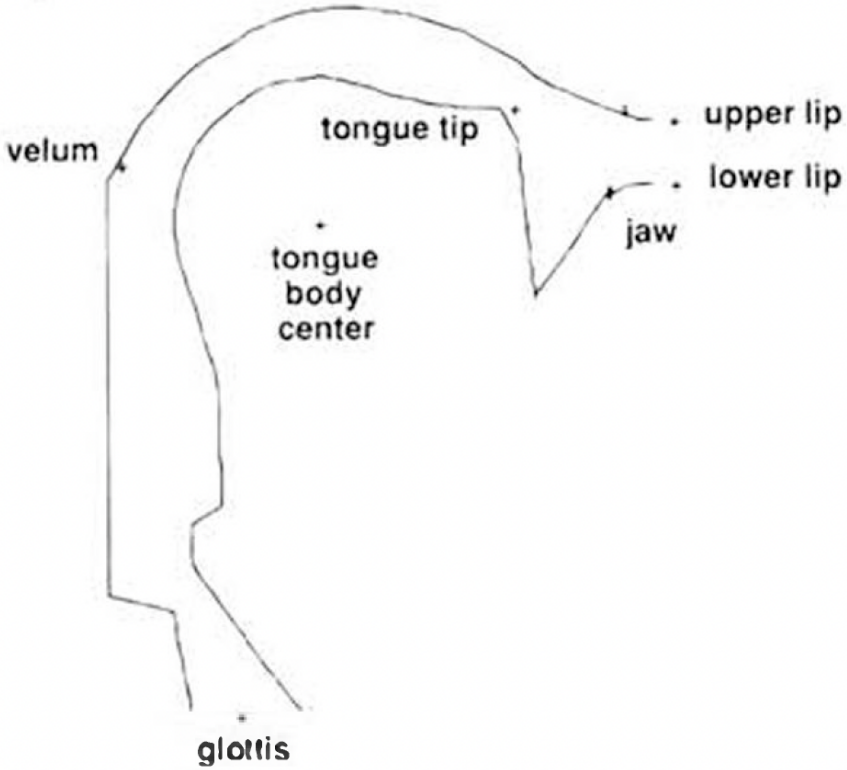
Outline

Tract variables are abstract

- Each tract variable is an abstract planning variable, not a single muscle
- Babies learn which tract variables are used to distinguish words
- Babies then learn how to coordinate their muscles to modify each tract variable

	tract variable	articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis





Outline

Tract variable dynamics are planned to be smooth

- Studies of planned movement (walking, arm reaching, and speech planning) show that trajectories are **much** smoother than is required by the muscle stiffness.
- Observed trajectories are well modeled by a cognitive-internal world model that plans continuous movements to optimally trade off
  - discrete requirements, e.g., gesture synchronization
  - continuous-valued criteria, e.g., smoothness and efficiency
- One such model (`https://www.haskinslaboratories.org/tada`) uses a nonlinear neural oscillator to plan each x_i:

$$\ddot{x}_i = -\alpha_i \dot{x}_i - \beta_i x_i^2 \dot{x}_i - \gamma_i \dot{x}_i^2 - \omega_{0i} x_i + \text{coupling terms}$$

Synchronization constraints

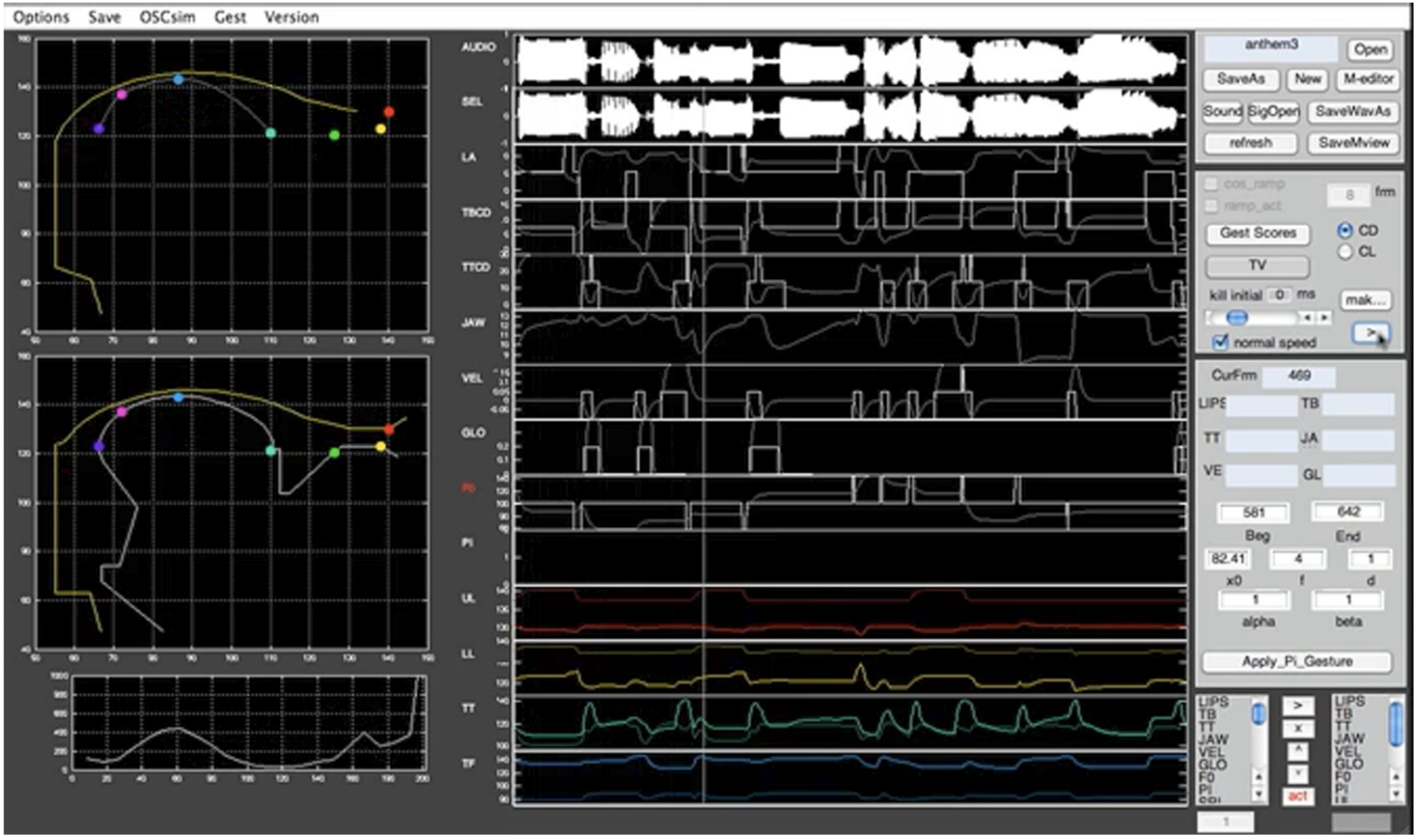
- Synchronization constraints are language-dependent
- In English CCVCC syllable,
  - solid line = gestures should be in-phase
  - dotted line = gestures should be 180° out of phase

Saltzman et al. (2008) A task-dynamic toolkit for modeling the effects of prosodic structure on articulation, Fig. 4

Tract variables drive muscles

- Once planned, the tract variable trajectories are converted by the brainstem into target lengths for individual muscles
  - In software, "muscles" are UL, LL, JAW, and so on
  - In a human, presumably, "muscles" are orbicularis oris, depressor anguli oris, levator anguli oris, and so on
- Muscle-internal local processing closely tracks each target length as a continuous function of time
  - In software, LA exactly matches its target
  - In a human, nerves within the muscle pump ATP into each fiber if its length is different from the desired length

TADA: Task Dynamic Articulatory synthesizer



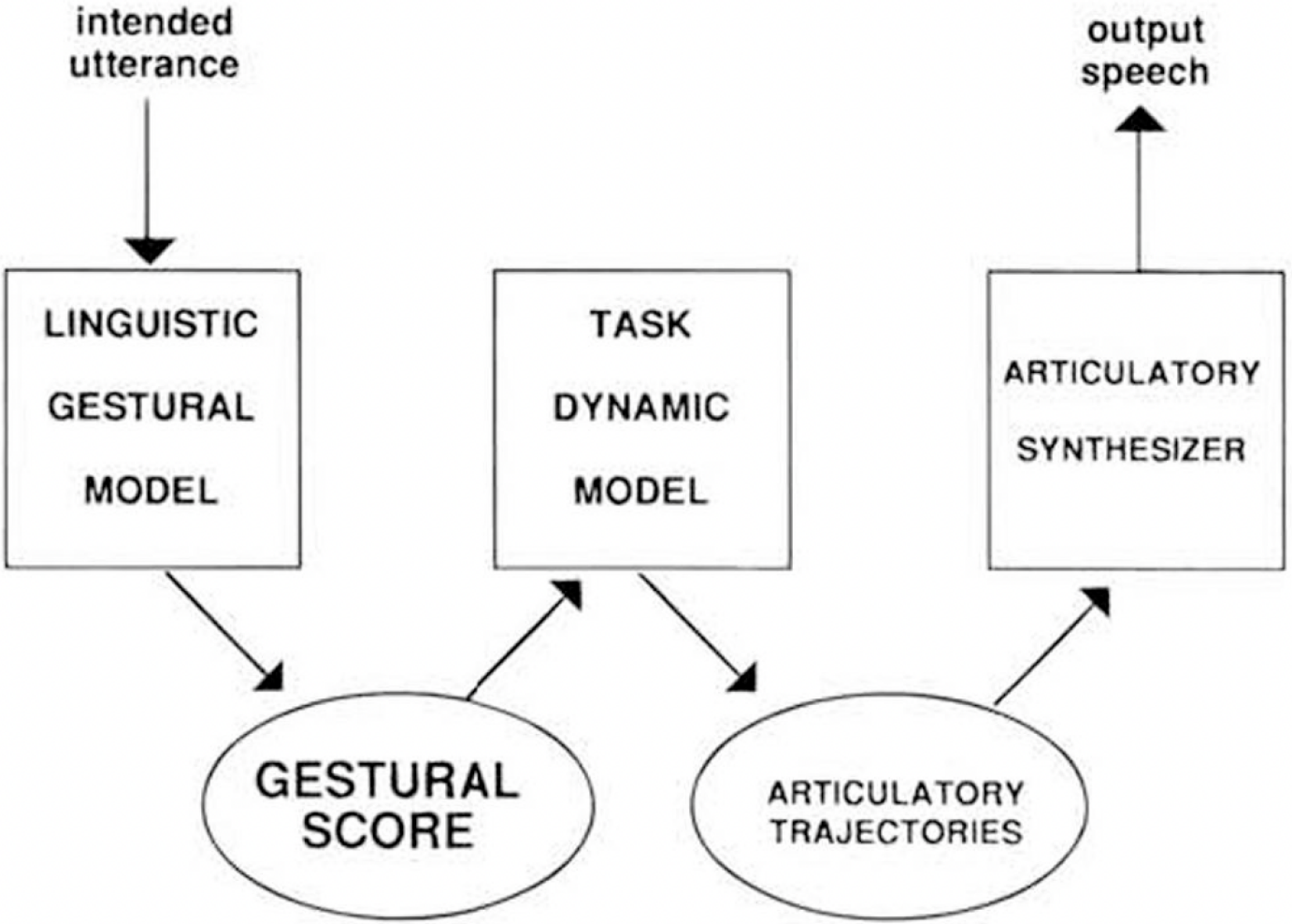
https://www.haskinslaboratories.org/tada

Outline

Example

Draw a gestural constellation, gestural score, and task dynamic implementation of the word “speech.”

Outline


intended utterance
LINGUISTIC GESTURAL MODEL
GESTURAL SCORE
TASK DYNAMIC MODEL
ARTICULATORY TRAJECTORIES
ARTICULATORY SYNTHESIZER
output speech