Embodying speech using computational models of the vocal tract

Presented to CHOC+UCI Pediatric Research and Data Science Group

August 3rd, 2023

Connor Mayer

Department of Language Science

University of California, Irvine

https://sites.socsci.uci.edu/~cjmayer/

Who am I?

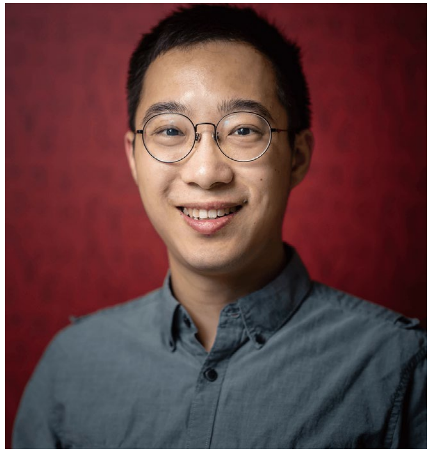
- Computational phonologist and phonetician
- Undergrad training in linguistics and computer science from the University of British Columbia
- Worked as a software developer on big budget video games for about 4 years
- Completed PhD in linguistics at UCLA in 2021
- Assistant professor at UCI since July 2021

My many collaborators

University of British Columbia

Bryan Gick

Yadong Liu



Arian Shamei

Gillian de Boer

Sid Fels

Jasia Azreen

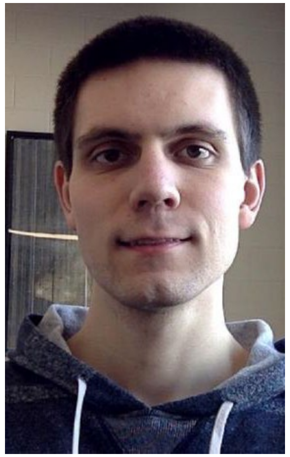
Ian Stavness
(U. Saskatchewan)

Chenhao Chiu
(National Taiwan U.)

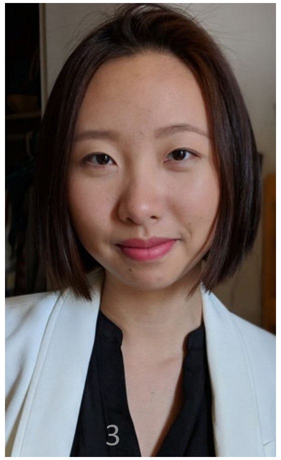
Monika Łuszczuk
(Maria Curie-Skłodowska U.)

Erik Widing

Francois
Roewer-Despres



Sophia Luo



What do linguists study?

Generally: Language as a cognitive faculty

More specifically:

- How do infants learn language so effectively/rapidly?
- Why do languages tend to have certain properties and not others?
- How and why do languages change over time?
- What kinds of mental representations do people have of their language(s)?

What I'll talk about today

Broadly speaking:

How do the **biomechanical structure** of our bodies and the **organization of speech motor control** shape speech?

Narrowly speaking:

Biomechanical simulation of speech movements can help us understand this better.

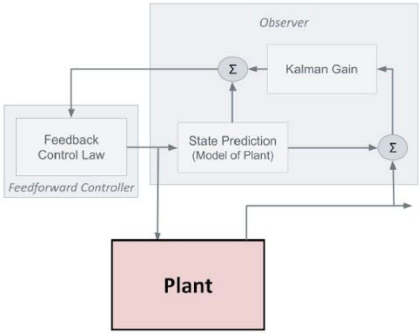
- Maybe some interesting applications in pediatrics?

Roadmap

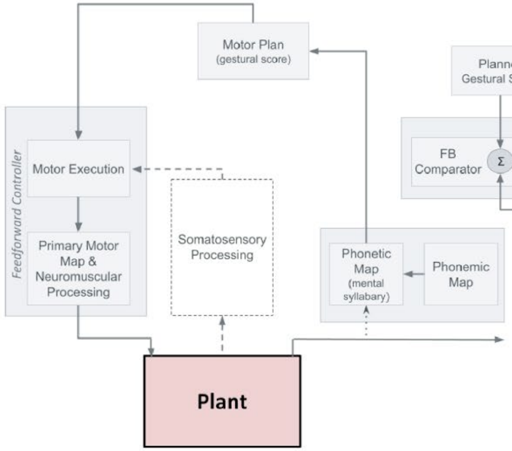
1. The body in models of speech motor control
2. The Artisynth simulation platform
3. Case study 1: Quantal lips
4. Case study 2: Lateral bracing
5. A (very preliminary) infant vocal tract model

Why is speech motor control interesting?

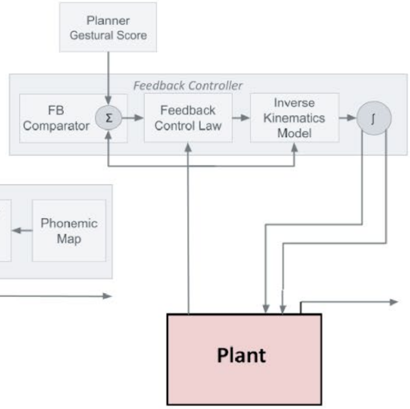
The body in models of speech motor control



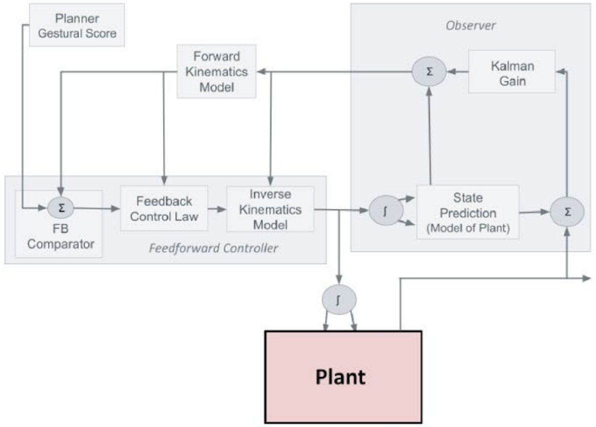
SFC
(e.g., Houde & Nagarajan 2011)



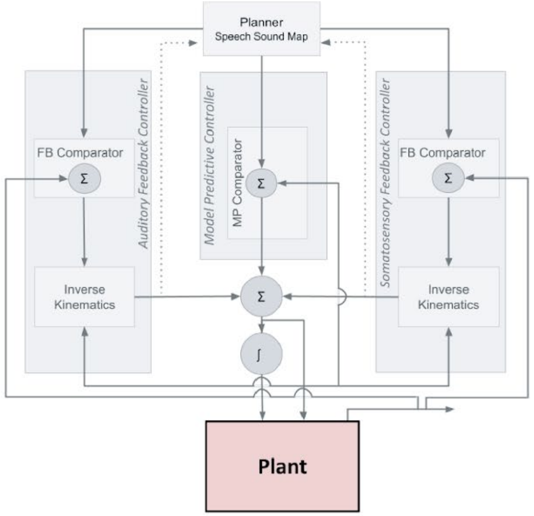
ACT
(e.g., Kröger 2009)



Task Dynamics
(e.g., Saltzman & Munhall 1989)



FACTS
(e.g., Parrell & al. 2019a)



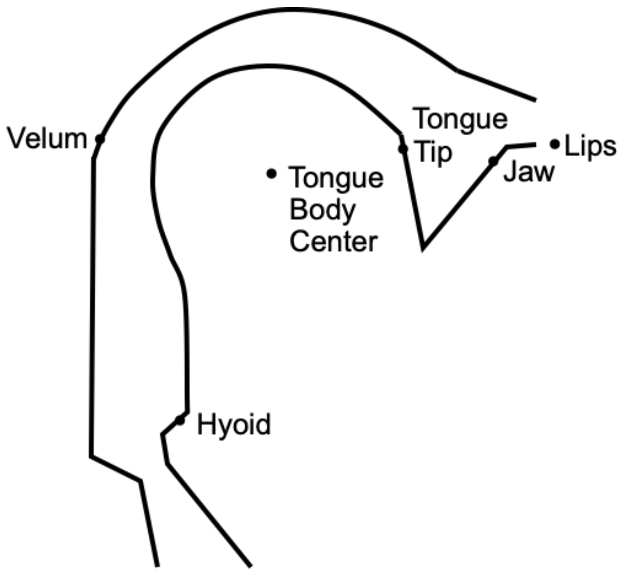
DIVA
(e.g., Guenther 2016)

Low-dimensional models of the vocal tract

Models of speech motor control typically use **low-dimensional representations** of the body

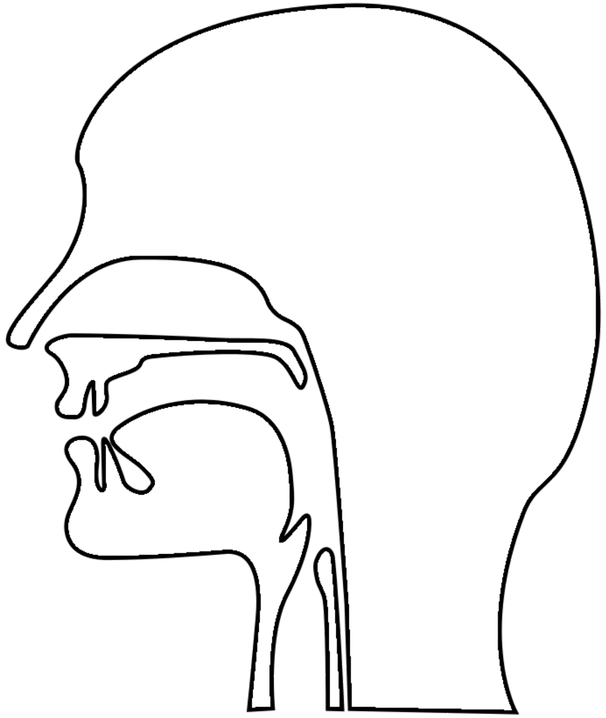
Parameters are **abstract articulators**

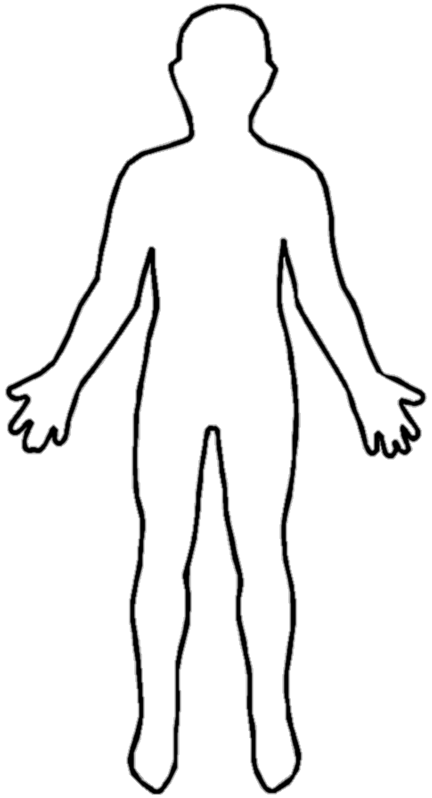
- “moving segments that have lengths but are massless...and are defined with reference to the simplified articulatory degrees of freedom [of the Mermelstein synthesizer]” (Kelso et al. 1986)



The Mermelstein synthesizer (Rubin et al. 1981)

Dimensionality reduction





Case study: What are the lips?

A single functional sphincter? (UCLA Phonetics Laboratory 2002)

A pair of independent structures? (Kelso et al. 1986, Guenther 2016)

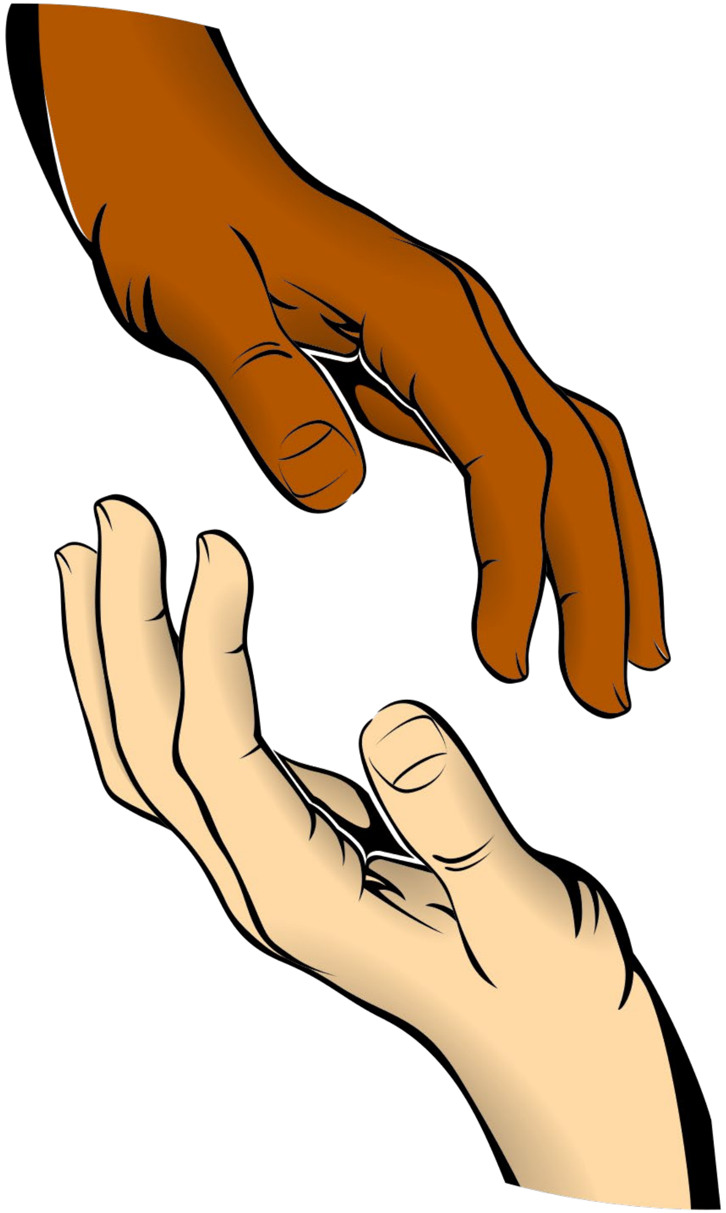
What are hands?

How does the hand move?

- Most of the muscles that move the hand are in the forearm

“Should those parts of the brain that regulate hand function be considered part of the hand? […] Although we understand what is meant conventionally by the simple anatomic term, we can no longer say with certainty where the hand itself, or its control or influence, begins or ends in the body.”

(Wilson 1998)

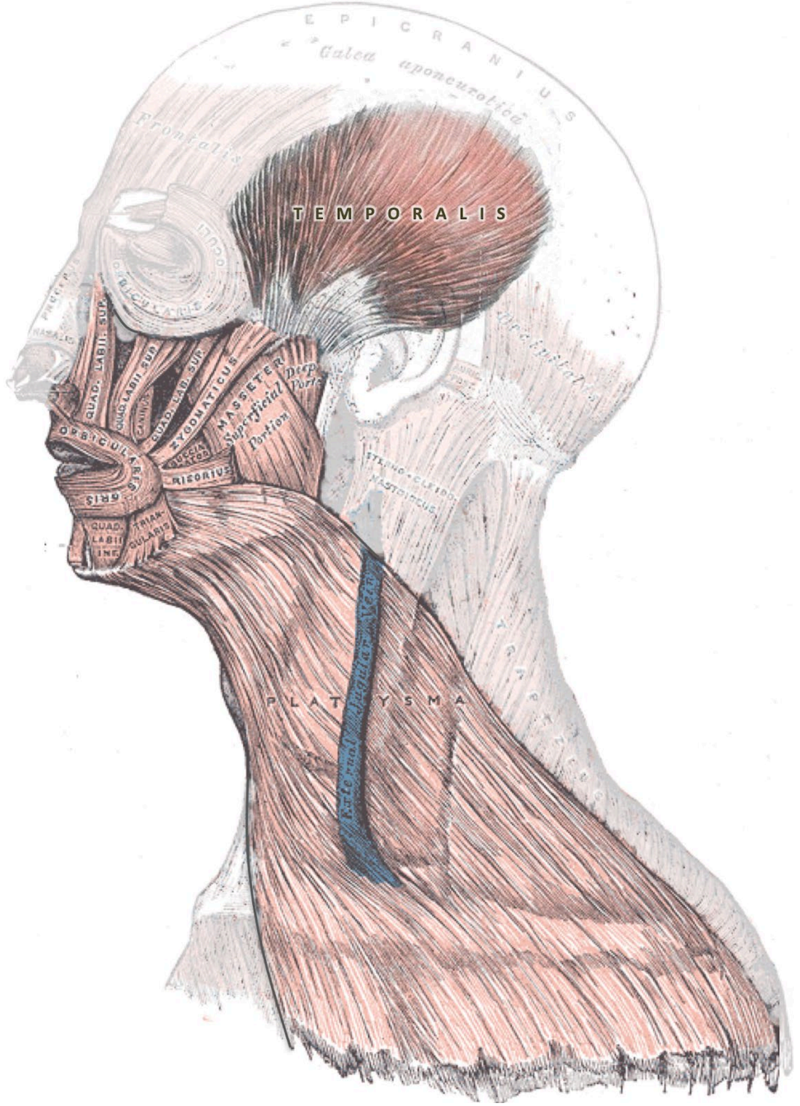


What are lips?

This is true of the lips as well!

- The muscles that affect lip position run from the forehead to the sternum

Defining the lips as a simple structure(s) **omits the anatomical and biomechanical properties** that underlie their function



(adapted from Gray and Lewis 1918, Plate 378)

Roadmap

1. The body in models of speech motor control
2. The Artisynth simulation platform
3. Case study 1: Quantal lips
4. Case study 2: Lateral bracing
5. A (very preliminary) infant vocal tract model

Artisynth: Starting from the body

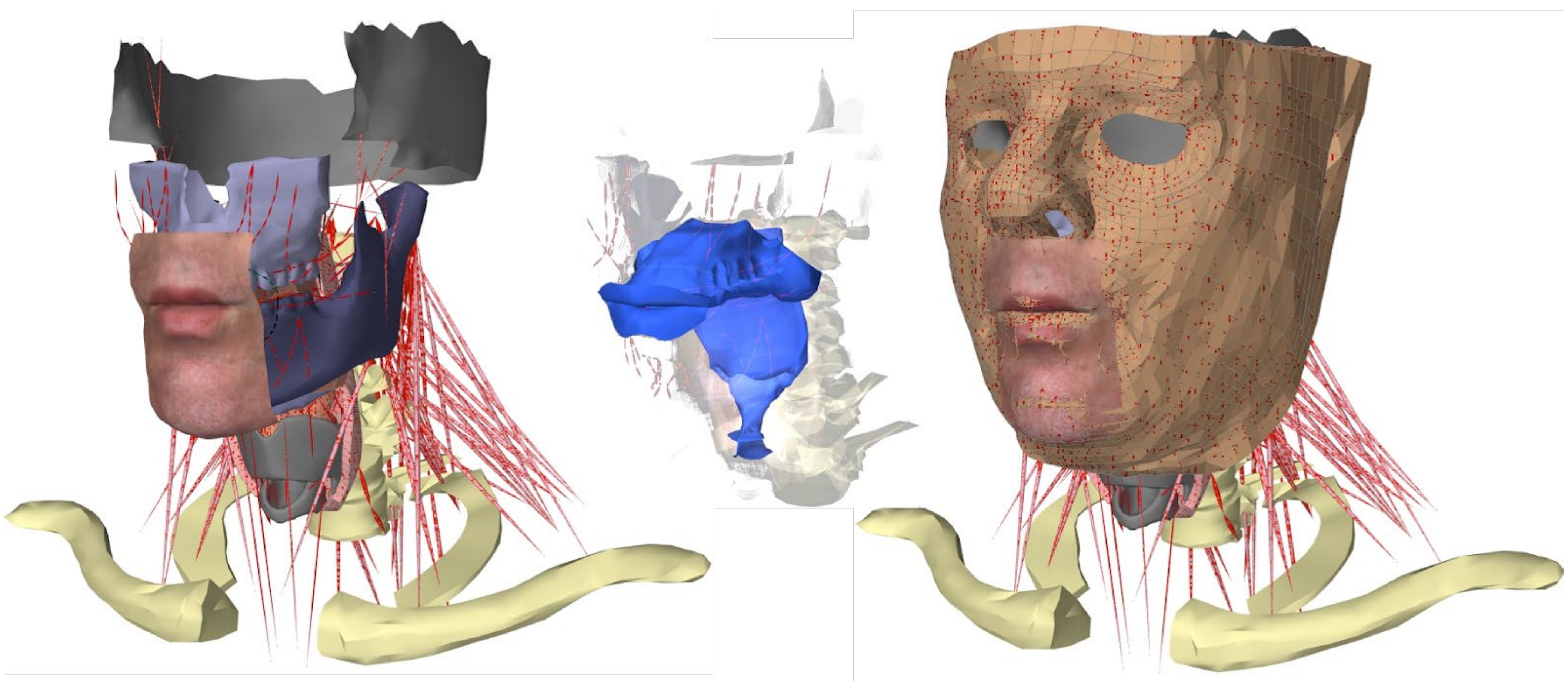
ArtiSynth is a biomechanical modeling platform (Stavness et al. 2012)

- Mixed multibody and FEM for rigid and deformable body structures
- FRANK: a biomechanical model of the head, neck, and vocal tract (Anderson et al. 2017)

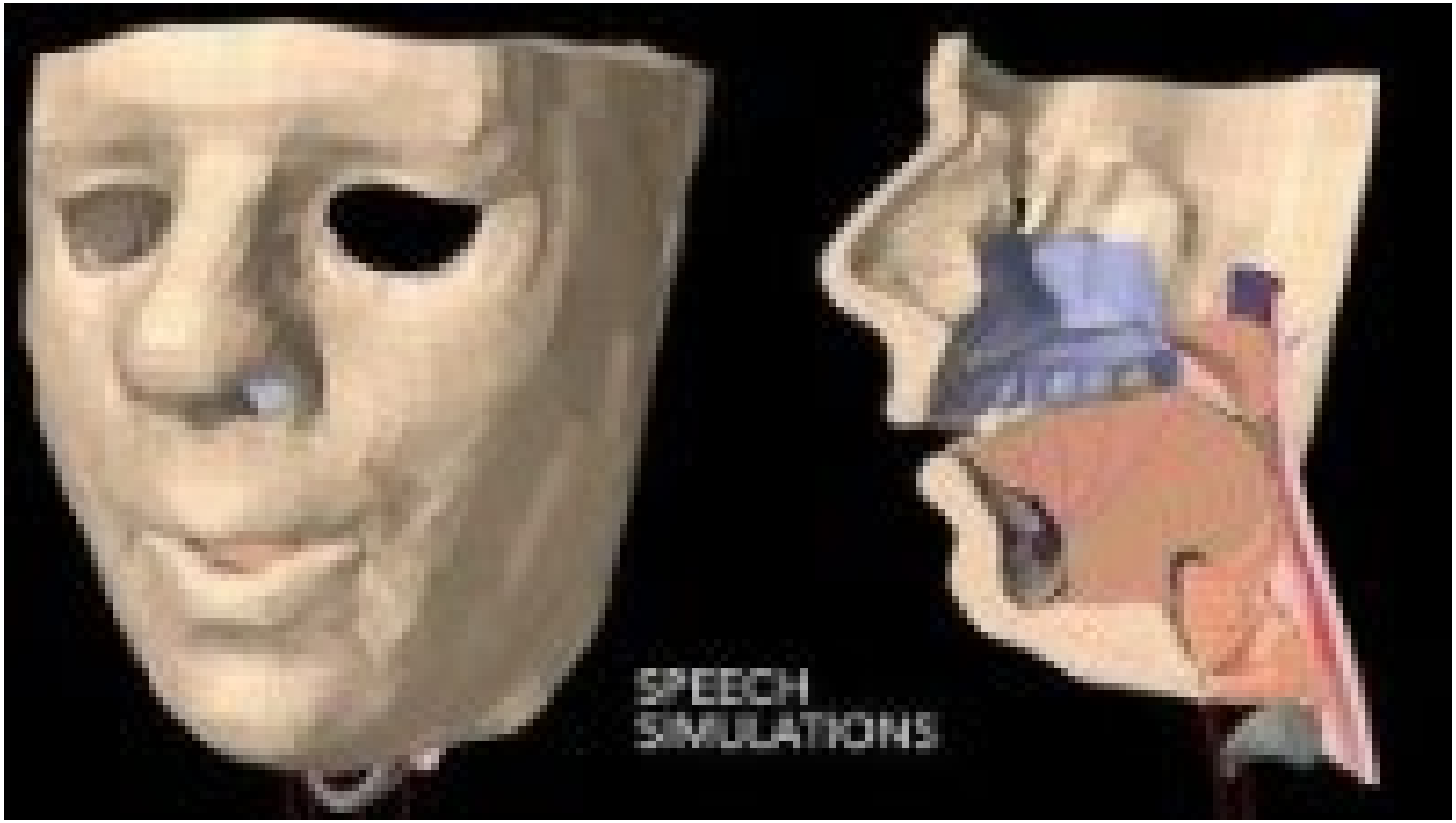
Biorealism has been a consistent priority in developing these models

- Based on medical imaging, fiber-level cryosections, specifications from existing literature
- Not designed with any theory of speech in mind

Artisynth: Starting from the body



Artisynth demo



Why use Artisynth?

It’s hard to collect data from speech muscles

- Muscles used in speech are difficult to access, highly interdigitated
- Artisynth predicts kinematics from muscle activations

Complements experimental work

Biomechanics are important for understanding neural control

- Relationship between muscle activation and movement is non-linear

Roadmap

1. The body in models of speech motor control
2. The Artisynth simulation platform
3. Case study 1: Quantal lips
4. Case study 2: Lateral bracing
5. A (very preliminary) infant vocal tract model

Case study 1: Quantal lips

General observation: Languages tend to use different lip shapes for different degrees of labial constriction.

Let's start by looking at the 451 languages in the UCLA Phonological Segment Inventory Database (UPSID; Maddieson 1984, Maddieson and Precoda 1990)

Gick, B., Mayer, C., Chiu, C., Widing, E., Roewer-Despres, F., Fels, S., and Stavness, I. (2020) Quantal biomechanical effects in speech postures of the lips. *Journal of Neurophysiology, 124*(3), 833 – 843.

UPSID labial typology (451 languages)

Though not without exceptions, there's a clear generalization:

- **Labial stops**: 99.8% bilabial (0.2% labiodental)
- **Labial fricatives**: 71% labiodental (29% bilabial)
- **Labial approximants**: 98% rounded (2% labiodental)

Why should this be the case?

A language could produce different degrees of constriction by **varying the activation of a single labial movement:**

- Labial stop: [p]
- Labial fricative: [p̞]
- Labial approximant: [p̞̞]

<u>Languages don’t do this!</u>

Why these mechanisms?

Mechanisms built for a task will be **robust to noisy, everyday conditions**
(e.g., Loeb 2012)

- Allow a **large margin of error**
- Optimize for **feed-forward function** (e.g., Perkell 2012; Guenther 2016)

Speech mechanisms with such properties are called **quantal**
(e.g., Stevens 1972; Stevens 1989; Stevens and Keyser 2010)

- **Large variation** in input → **little response** in output

Quantal effects in speech

Speech exploits ***quantal*** properties of the vocal tract

<u>Acoustics</u>

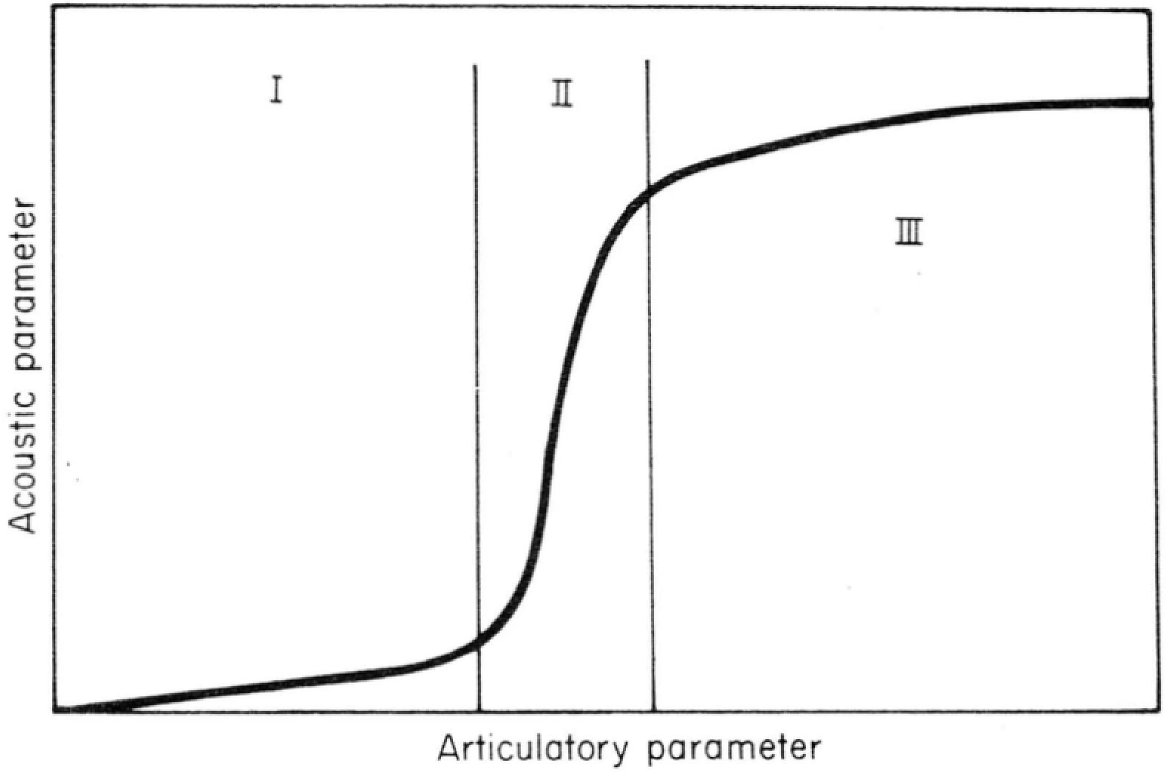
Large Δ in articulator position

→ minimal Δ in acoustics (Stevens 1989)

<u>Biomechanics</u>

Large Δ in muscle activation

→ small Δ in articulator position



Past work on quantal biomechanics

Limited discussion of quantal biomechanical effects
(e.g., Fujimura and Kakita 1979; Fujimura 1989; Perkell et al. 2004; Perkell 2012)

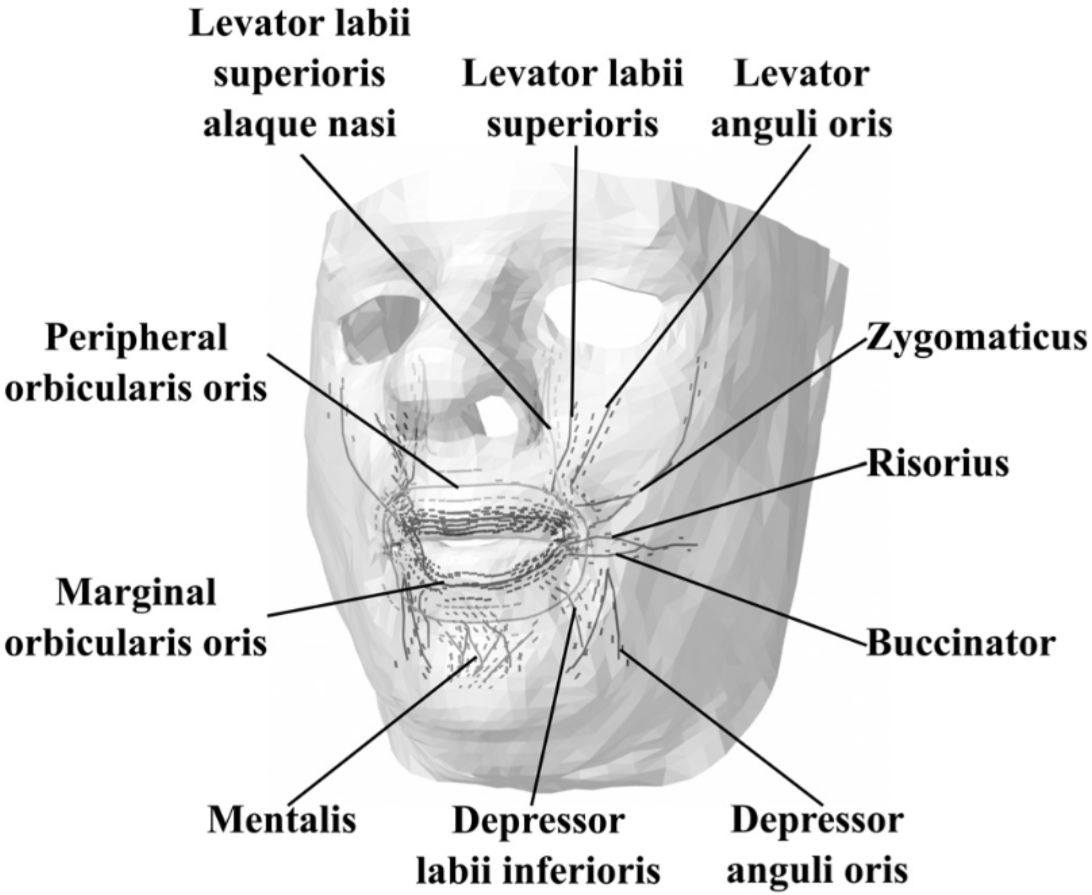
Simulation studies have demonstrated quantal effects in

- The soft palate (Gick et al. 2014; Anderson et al. 2019)
- The larynx (Moisik and Gick 2017)
- Lip rounding with variations in muscle stiffness (Nazari et al. 2011)

<u>Not all sets of muscle activations exhibit quantality!</u>
(Gick et al. 2014; Moisik and Gick 2017)

The current study

Tests for quantal effects in the three canonical lip postures using Artisynth



Assumptions

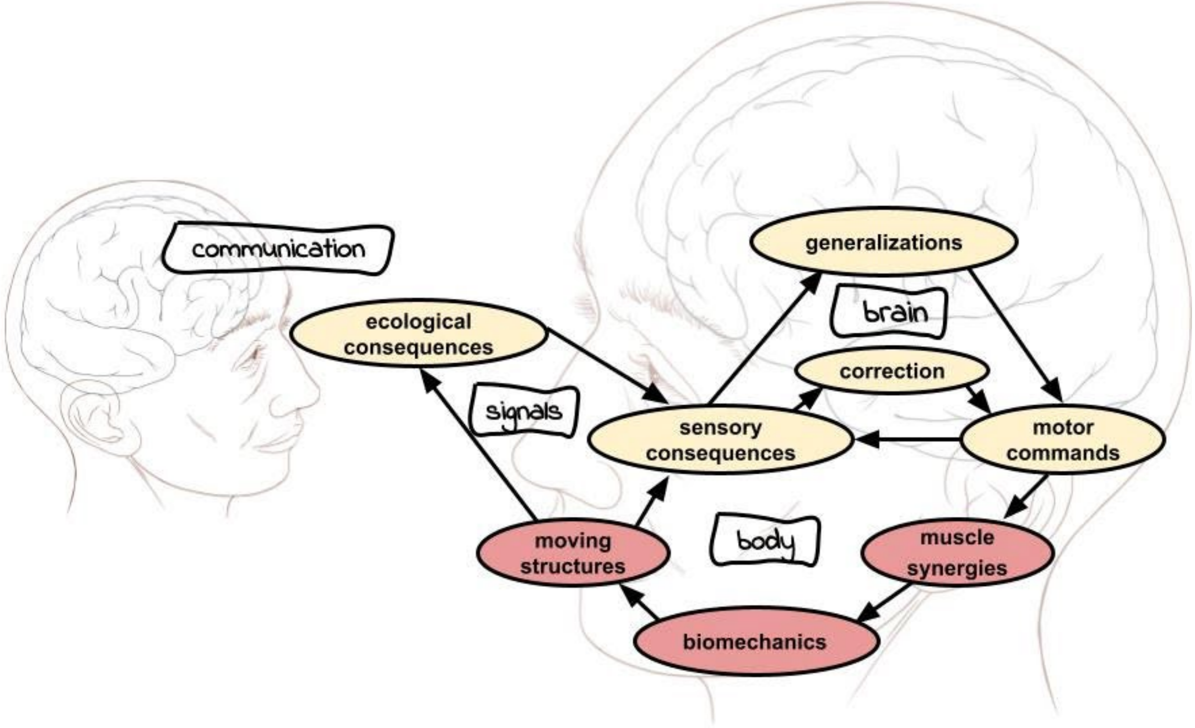
- Speech movements are generated by functionally independent groupings of muscles that activate in fixed proportion (**modules**)
  (e.g., Bernstein 1967; Ting et al. 2015)

- Selected in part based on **robustness**

Assumptions: Modular control

The ‘body parts’ in our models should be defined **functionally**

- We’ve adopted the notion of motor **modules**
- (e.g. Bernstein 1967, Fowler & Turvey 1978, Ting et al. 2015, d’Avella et al. 2015)
- A **body-based structure that can reliably generate a phonetic movement and its communicatively relevant sensory consequences**



Predictions

Canonical lip modules will be

1. **Robust** across a wide range of **activation levels**
2. **Robust to interference** from surrounding muscles

Simulation 1: Robustness to varying activation

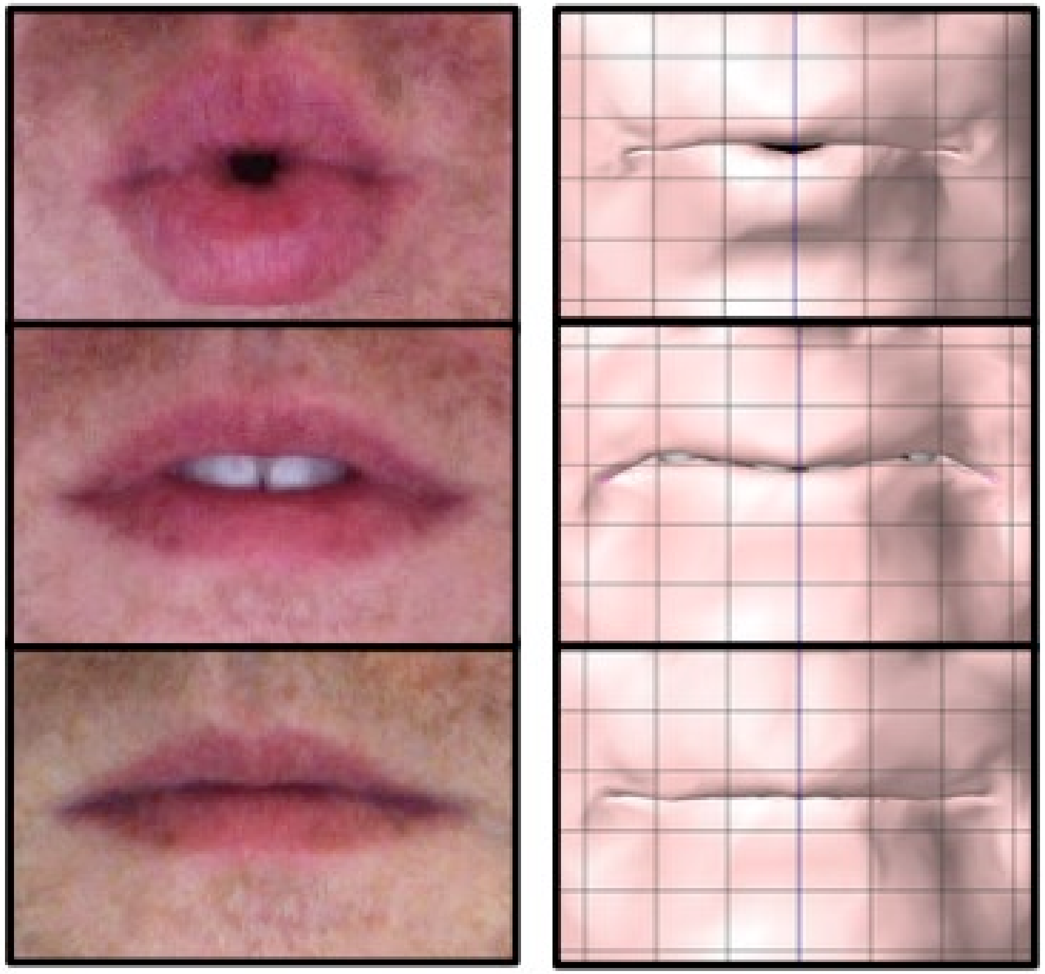
- Defined muscle groupings based on known muscle involvements
  (Lightoller 1925; Stavness et al. 2013)

- No “right” choice: many inputs will contain the necessary mechanic
  (e.g., Loeb 2012)

	OOPs	OOPi	OOMs	OOMi	MENT	RIS	LLSAN	LLS
Bilabial	–	–	30	30	20	20	–	–
Labiodental	–	–	–	26	26	26	36	50
Rounded	40	40	–	–	–	–	–	–

Simulation 1: Robustness to varying activation

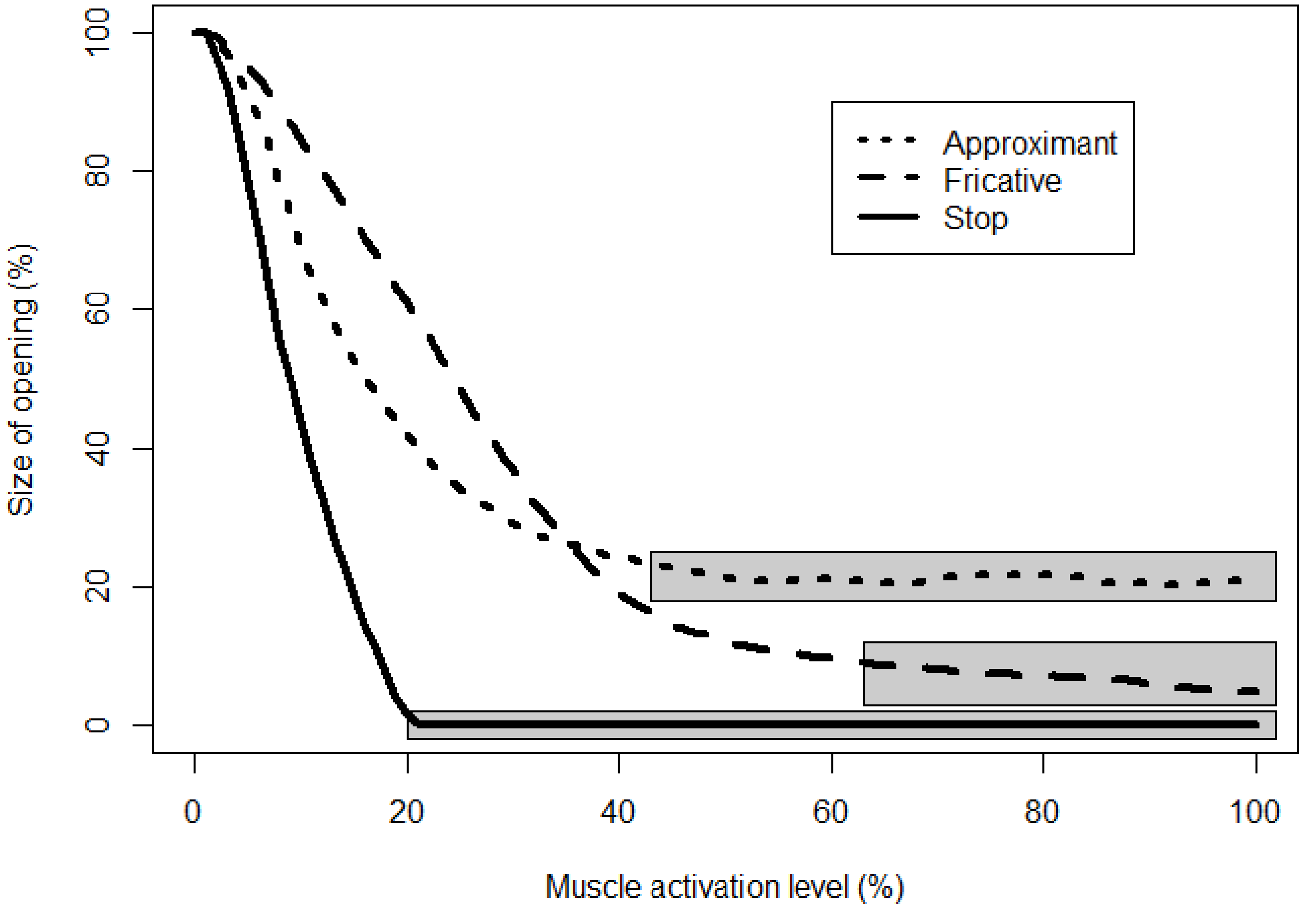
- Activated muscle groupings up to maximum stresses
- Measured opening size at different activation levels



Simulation 1: Results

Non-linearities occur as predicted!

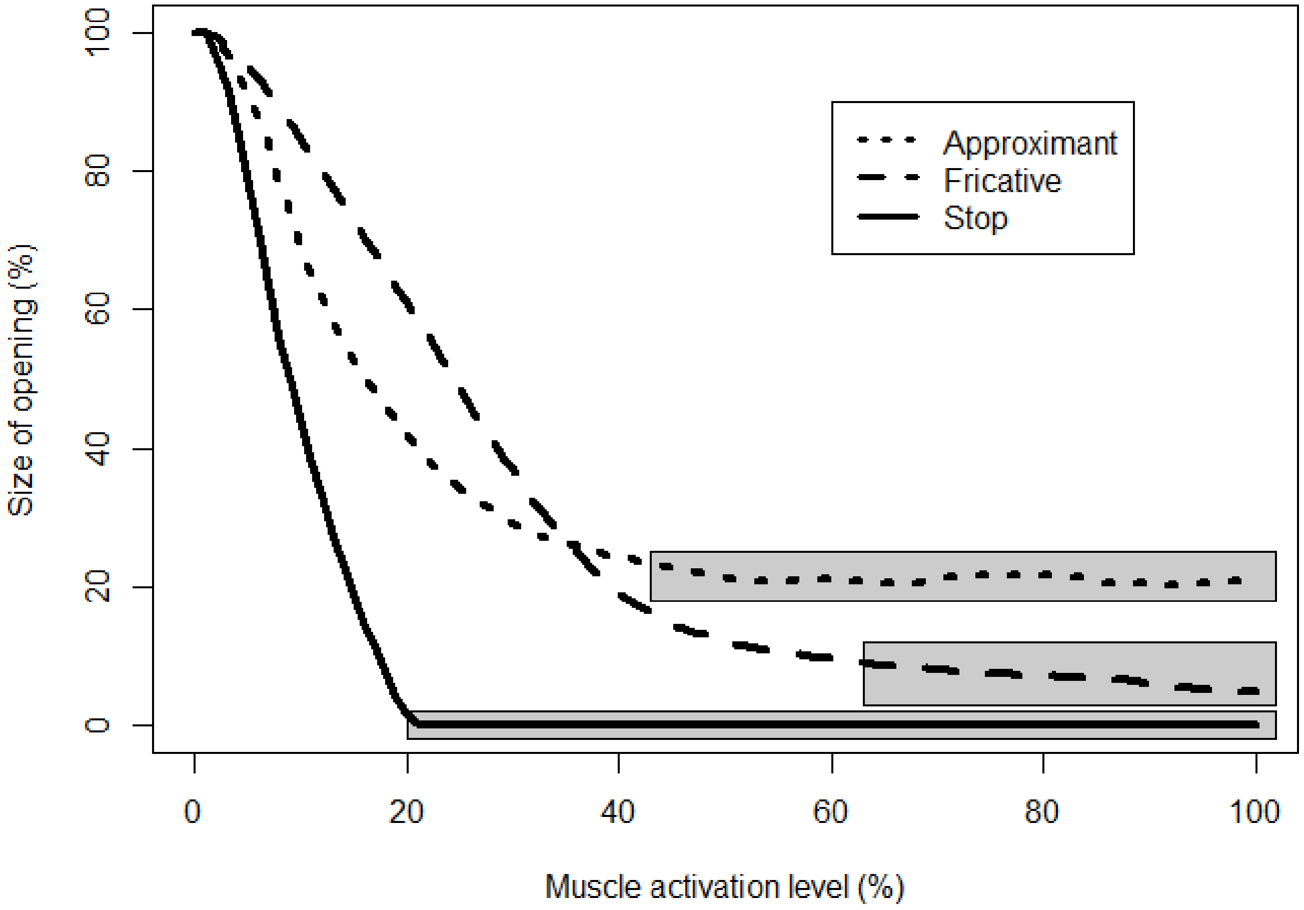
- Grey boxes: areas where 95% of distance to maximum closure has been covered



Simulation 1: Results

Takeaway

All three speech postures are <u>robust to variation in activation levels</u> of relevant muscle groups



Discussion

Why don’t we see labial inventories that look [p], [p̞], [p̞̞]?

- The regions in which frication and approximation are achievable using this configuration are <u>biomechanically unstable</u>.

The sets of muscles associated with the three canonical lip postures are:

1. Robust to <u>intrinsic</u> activation noise (Simulation 1)
2. Robust to <u>extrinsic</u> noise from surrounding muscles (Simulation 2)

Mayer et al. (2021) show this for two additional lip postures

Mayer, C., Chiu, C., & Gick, B. (2021). Biomechanical simulation of lip compression and spreading. *Canadian Acoustics, 49*(3), 38-39.

Discussion

Bears on theories of speech organization and motor control

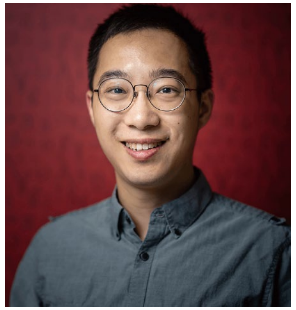
- Degree of constriction and involved articulators are <u>not independent parameters!</u>
- Primitive units of organization are **modular muscle groupings** that **activate in a fixed proportion** to achieve a **particular functional goal**
  (e.g., Bernstein 1967; Safavynia and Ting 2013; Gick and Stavness 2013; Ting et al. 2015)

Understanding these structures provides explanatory power for linguistic phenomena.

Roadmap

1. The body in models of speech motor control
2. The Artisynth simulation platform
3. Case study 1: Quantal lips
4. Case study 2: Lateral bracing
5. A (very preliminary) infant vocal tract model

Case Study 2: Liu et al. (2022b)



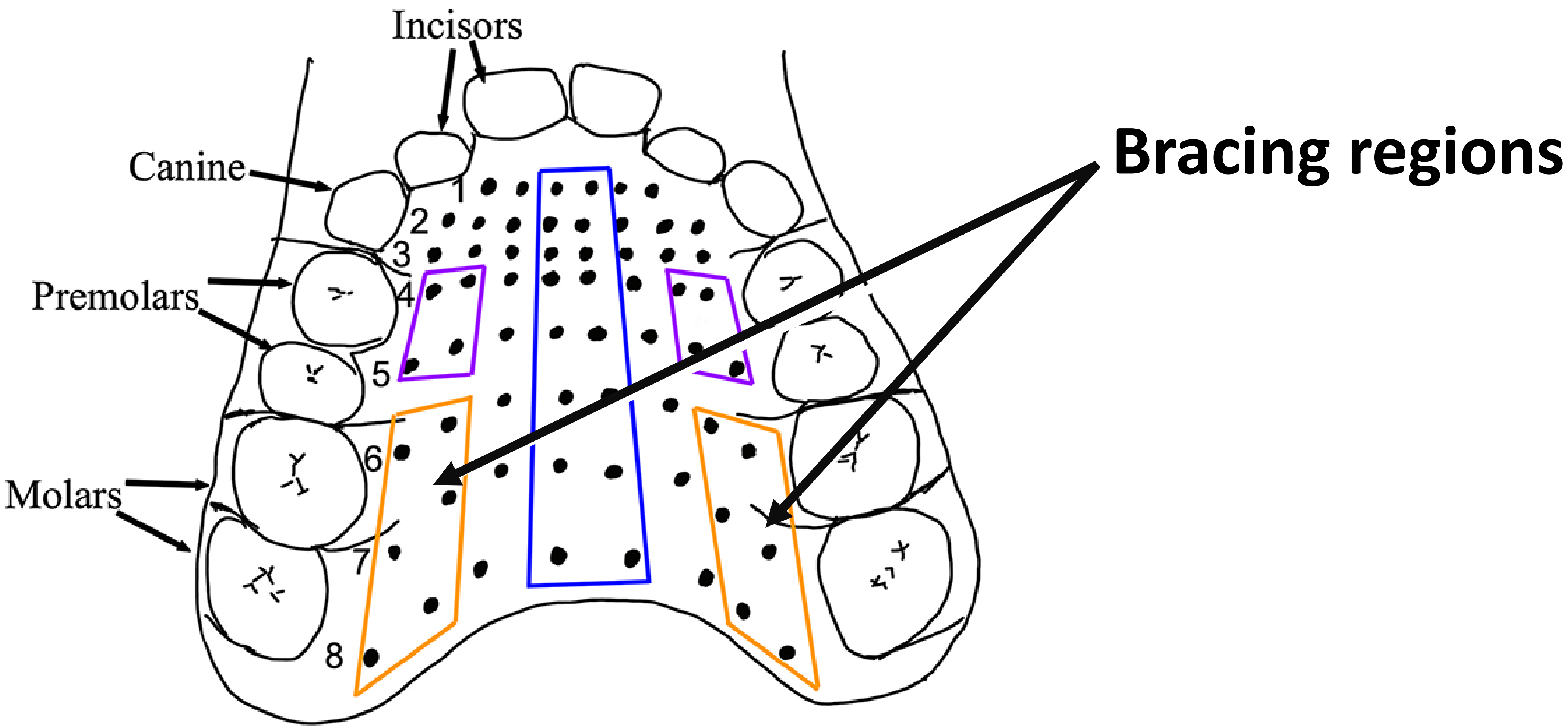
Lateral bracing: sides of tongue in contact with hard palate/upper molars

- Similar to oral preparatory phase in swallowing (Mayer et al. 2017)
- Separates central oral tract from buccal cavities (Perkell 1979, a.o.)

Also used pervasively in speech:
- Forms closed aeroacoustic tube (Gick et al. 2017)
- Facilitates movement of tongue (Stone 1990)
- Provides somatosensory feedback (Stevens and Perkell 1977)

Liu, Y., Luo, S., Łuszczuk, M., Mayer, C., Shamei, A., de Boer, G., & Gick, B. (2022). Robustness of lateral tongue bracing under bite block perturbation. *Phonetica, 79*(6), 523-549.

Lateral bracing



Lateral bracing in speech

Lateral bracing is maintained almost constantly during speech
(Gick et al. 2017, Liu et al. 2022a,b)

- Exceptions are lateral sounds like /l/ and some low vowels like /ɑ/

Liu et al. (2022a) propose that bracing be treated as a speech **posture**

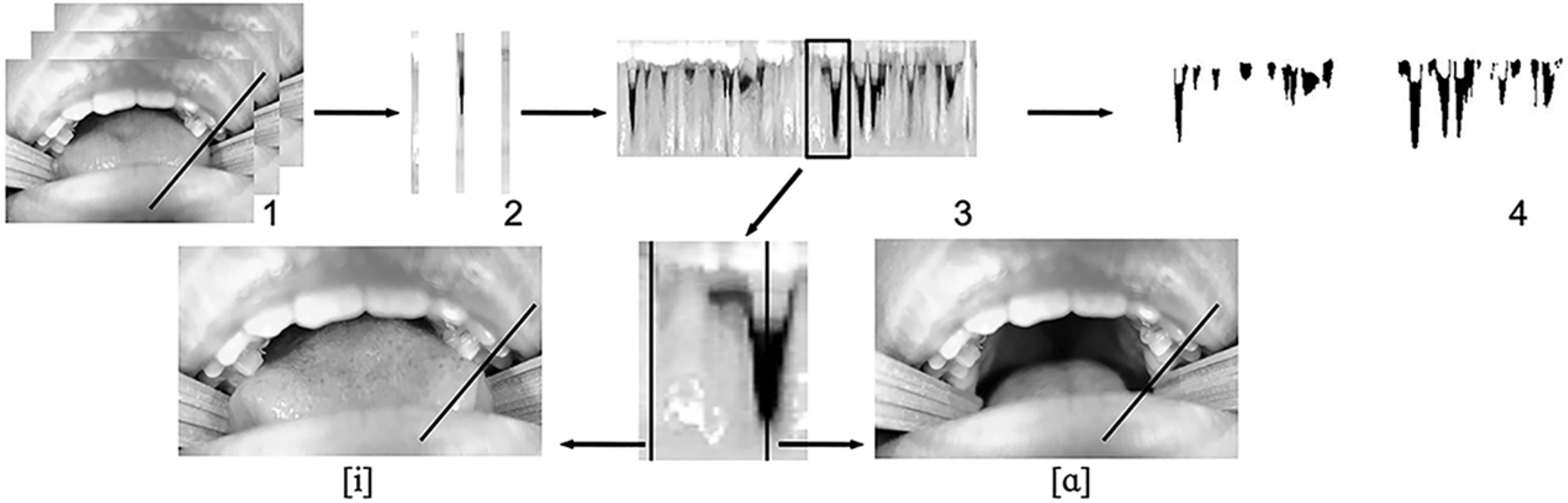
- A substratum of tonic muscle activation underlying other movements

Question: Other body postures exhibit robustness to perturbation. Is lateral bracing robust in the same way?

Part 1: Experimental study

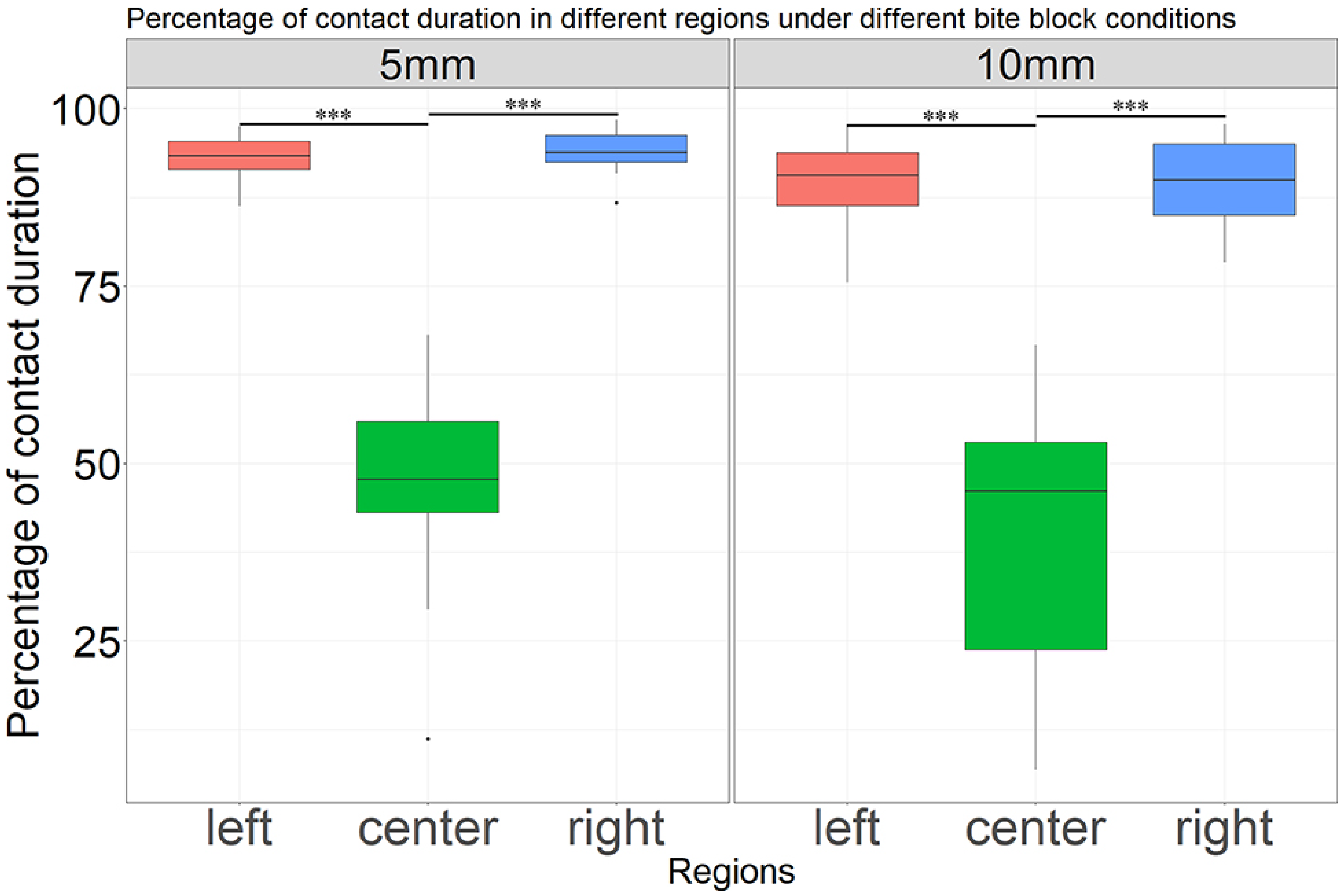
Perturbed speakers using 5mm and 10mm bite blocks between teeth

Bracing measured using camera



(a) Videokymogram illustration and two corresponding frames in an enlarged section

Results


Percentage of contact duration in different regions under different bite block conditions
5mm
10mm
100
75
50
25
***
***
***
***
Percentage of contact duration
left
center
right
left
center
right
Regions

Part 2: Simulation study

Question: does maintaining lateral bracing under perturbation require active effort?

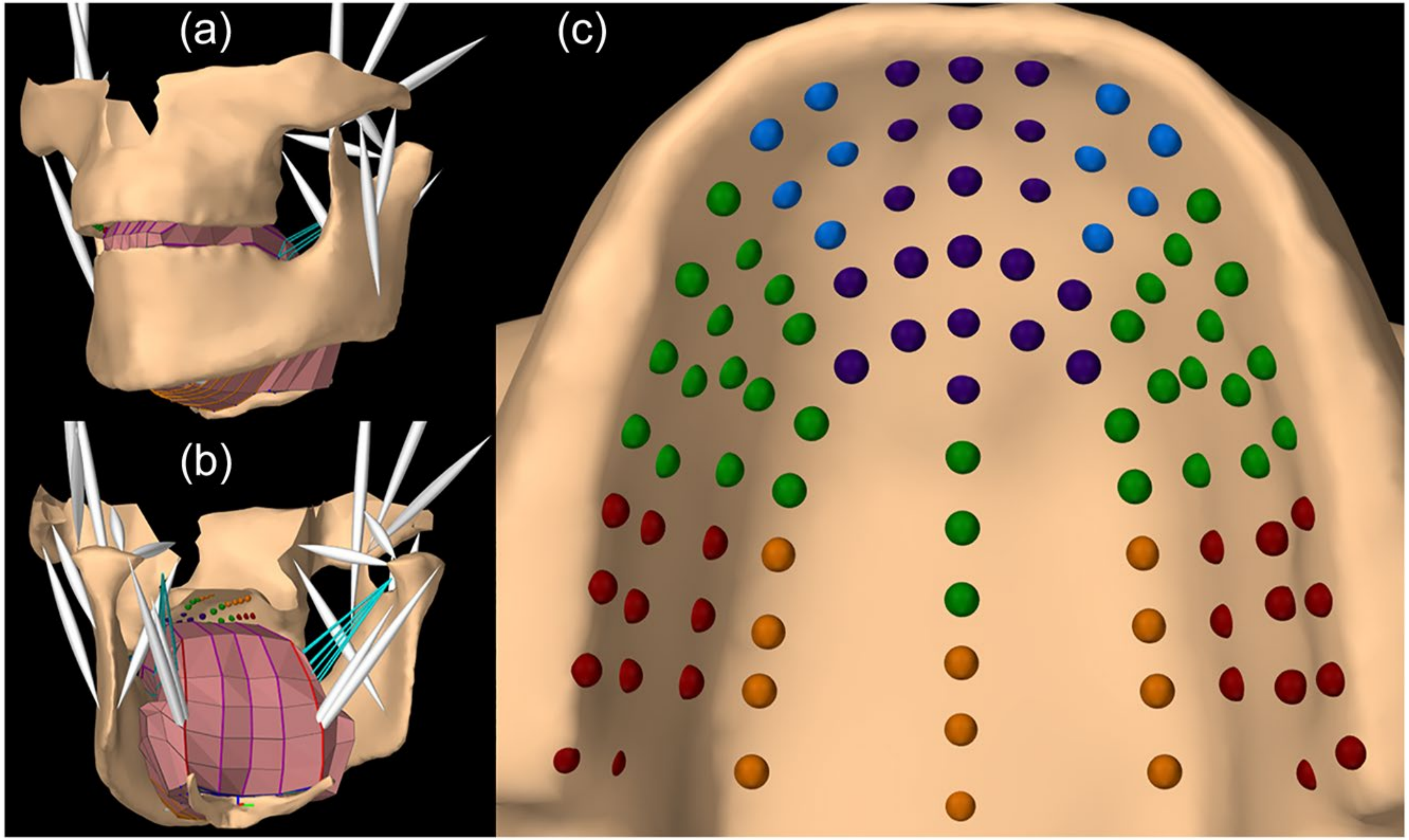
Simulation study:

- Activated 10 intrinsic and extrinsic tongue muscles at all combinations of 3 different activation levels (3^{10} = ~70k simulations)

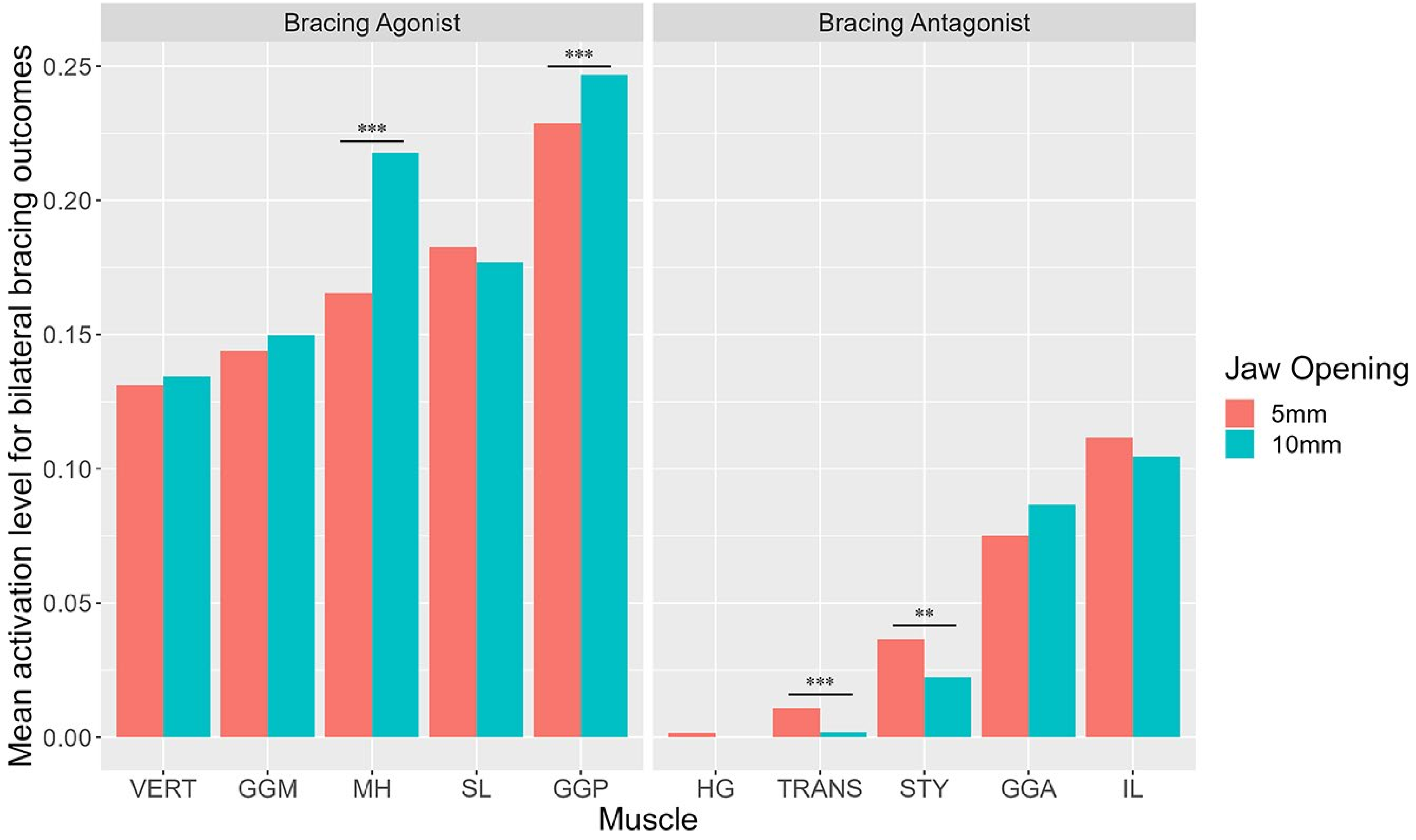
- Two conditions: 5mm vs. 10mm jaw aperture

- Detected bracing outcomes using virtual electropalatogram

Virtual electropalatogram



Results



Bracing Agonist
Bracing Antagonist
Mean activation level for bilateral bracing outcomes
0.25
0.20
0.15
0.10
0.05
0.00
***
***
***
**
VERT
GGM
MH
SL
GGP
HG
TRANS
STY
GGA
IL
Muscle
Jaw Opening
5mm
10mm

Discussion

Lateral bracing is actively maintained during speech

- Behaves like other body postures

Future directions:

- How unilateral bracing is achieved (Azreen et al. in press)
- Interplay between postural and transient movements
- How important is bracing for effective tongue control?

Roadmap

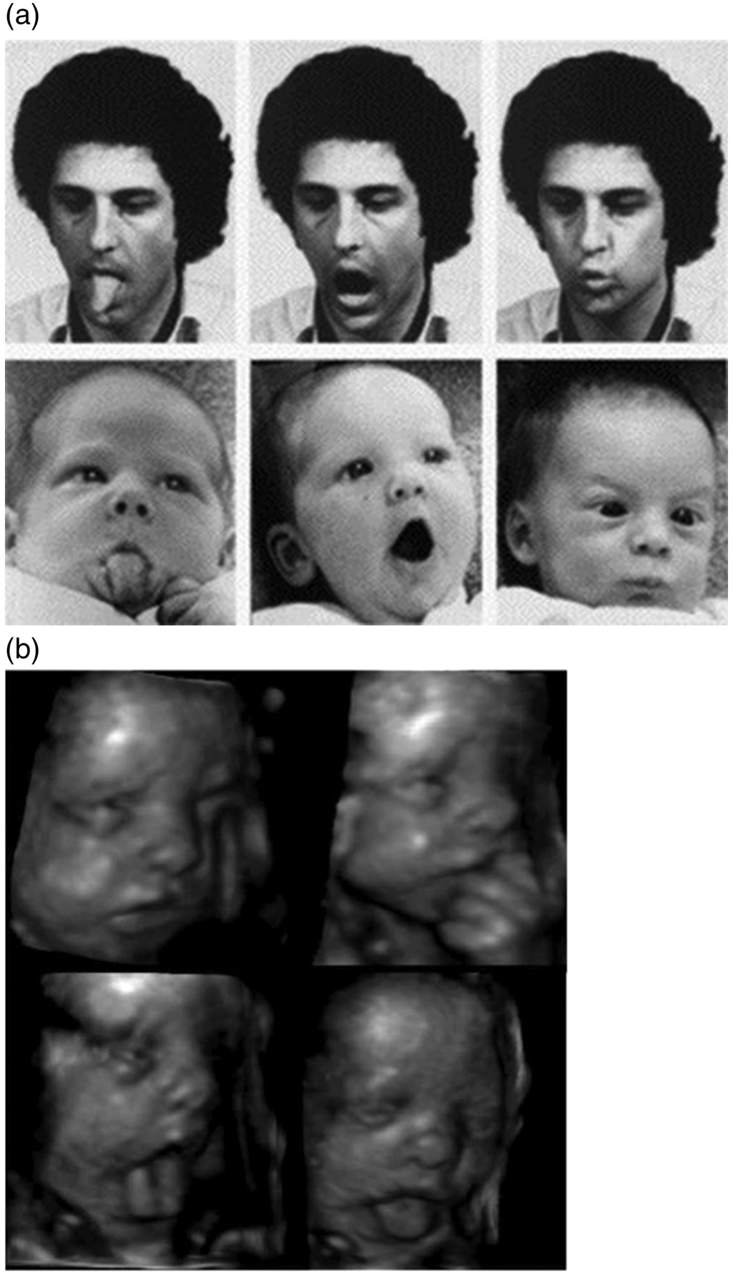
1. The body in models of speech motor control
2. The Artisynth simulation platform
3. Case study 1: Quantal lips
4. Case study 2: Lateral bracing
5. A (very preliminary) infant vocal tract model

Bootstrapping speech

Keven and Akins (2017) propose that infant tongue protrusion is an innate motor behavior, not imitation.

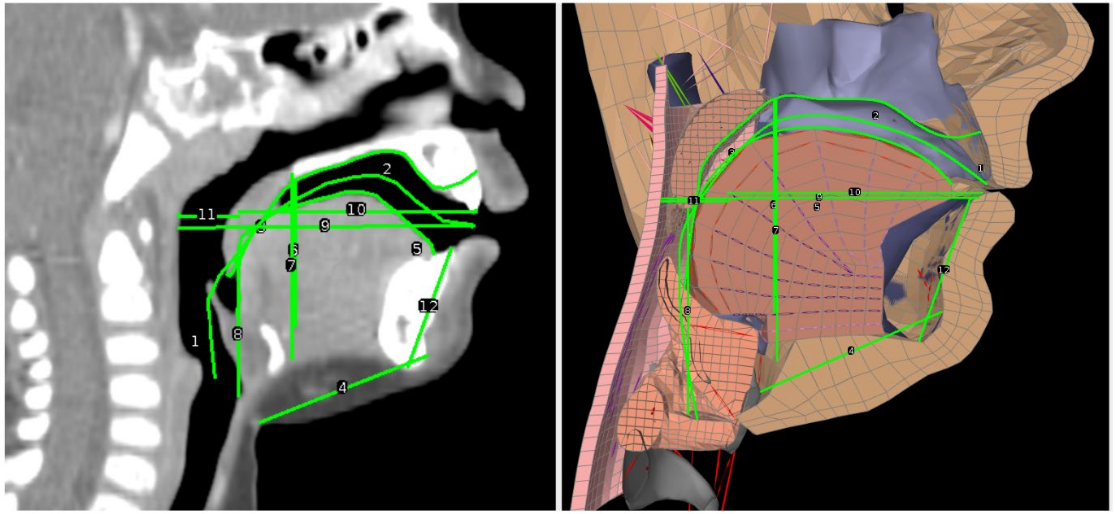
- Serves as a starting point for 'bootstrapping' other motor behaviors.

Mayer et al. (2017) propose that speech movements may be bootstrapped from aerodigestive movements.



Mayer, C., Roewer-Despres, F., Stavness, I. & Gick, B. (2017). Do innate stereotypies serve as a basis for swallowing and learned speech movements? *Behavioral and Brain Sciences, 40.* doi:10.1017/S0140525X16001928

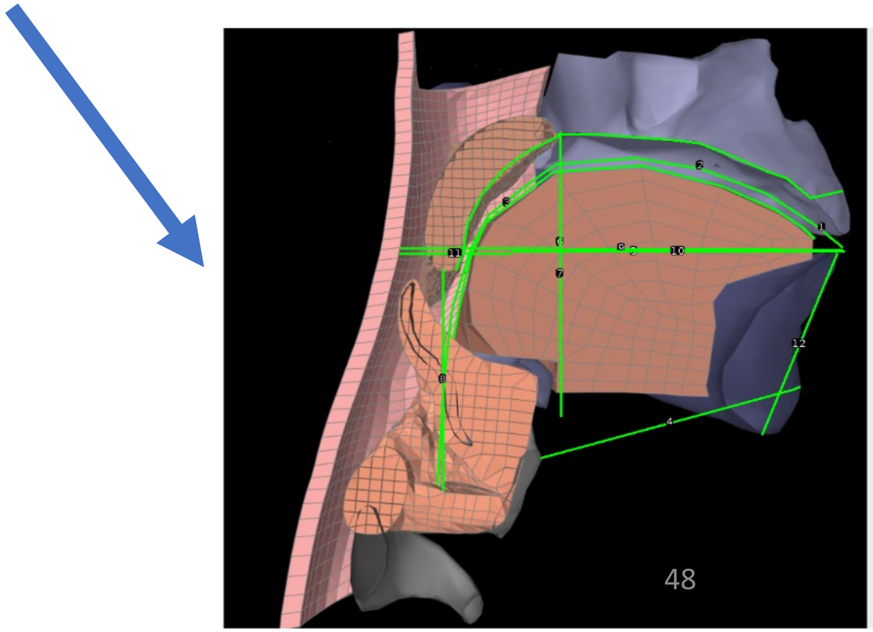
A (very preliminary) infant vocal tract model



Mayer et al. (2018) created a simple infant vocal tract model by warping adult model

Want to make a model based on infant imaging data, etc.

- Hard to come across for healthy populations!



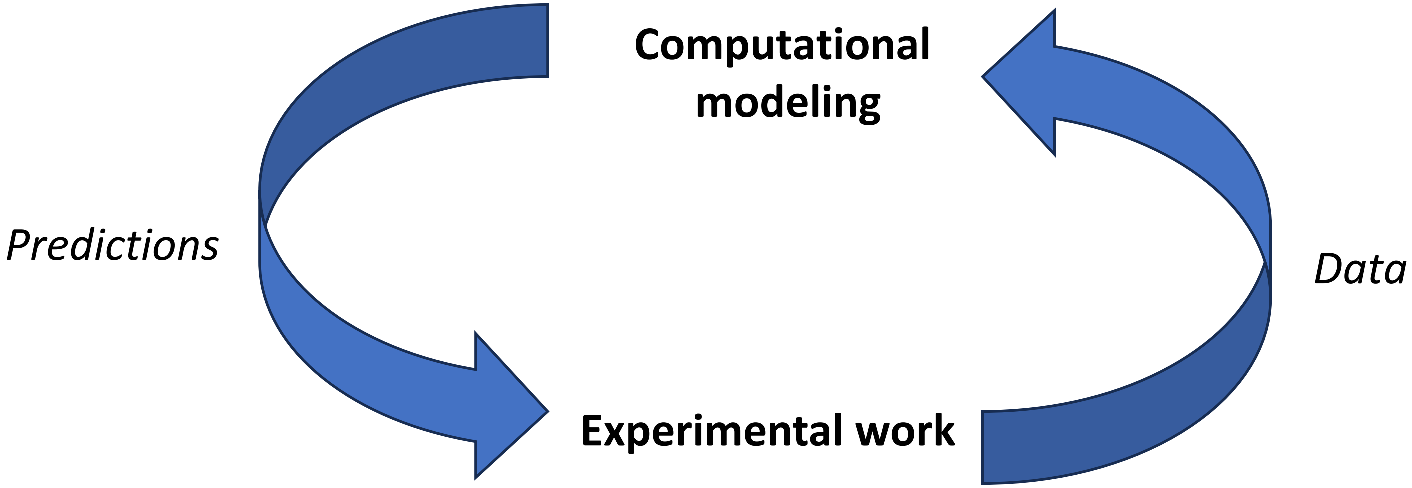
Mayer, C., Stavness, I., & Gick, B. (2018). A biomechanical model for infant speech and aerodigestive movements. *Canadian Acoustics, 46*(4), 30-31.

Conclusion

Models of the vocal tract can provide insight into

- how speech movements are controlled
- how biomechanics shapes speech

Goal: ‘close the loop’ between experimental and computational work

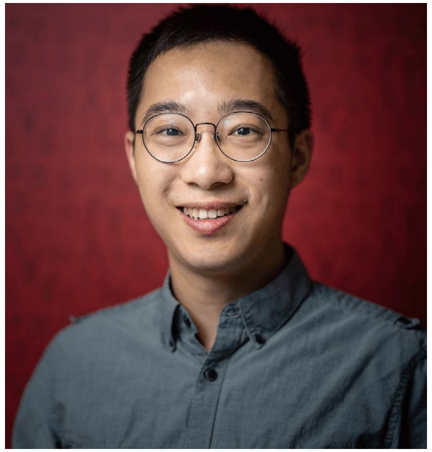


My many collaborators

University of British Columbia

Bryan Gick

Yadong Liu



Arian Shamei

Gillian de Boer

Sid Fels

Jasia Azreen

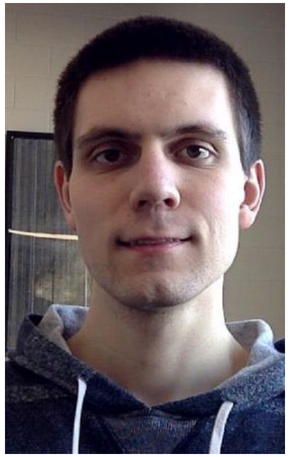
Ian Stavness
(U. Saskatchewan)

Chenhao Chiu
(National Taiwan U.)

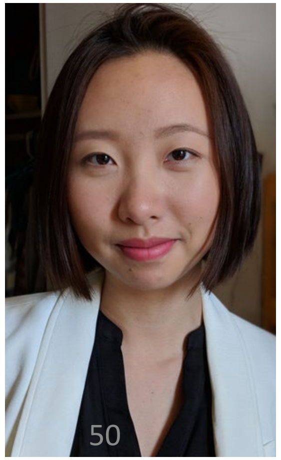
Monika Łuszczuk
(Maria Curie-Skłodowska U.)

Erik Widing

Francois
Roewer-Despres



Sophia Luo

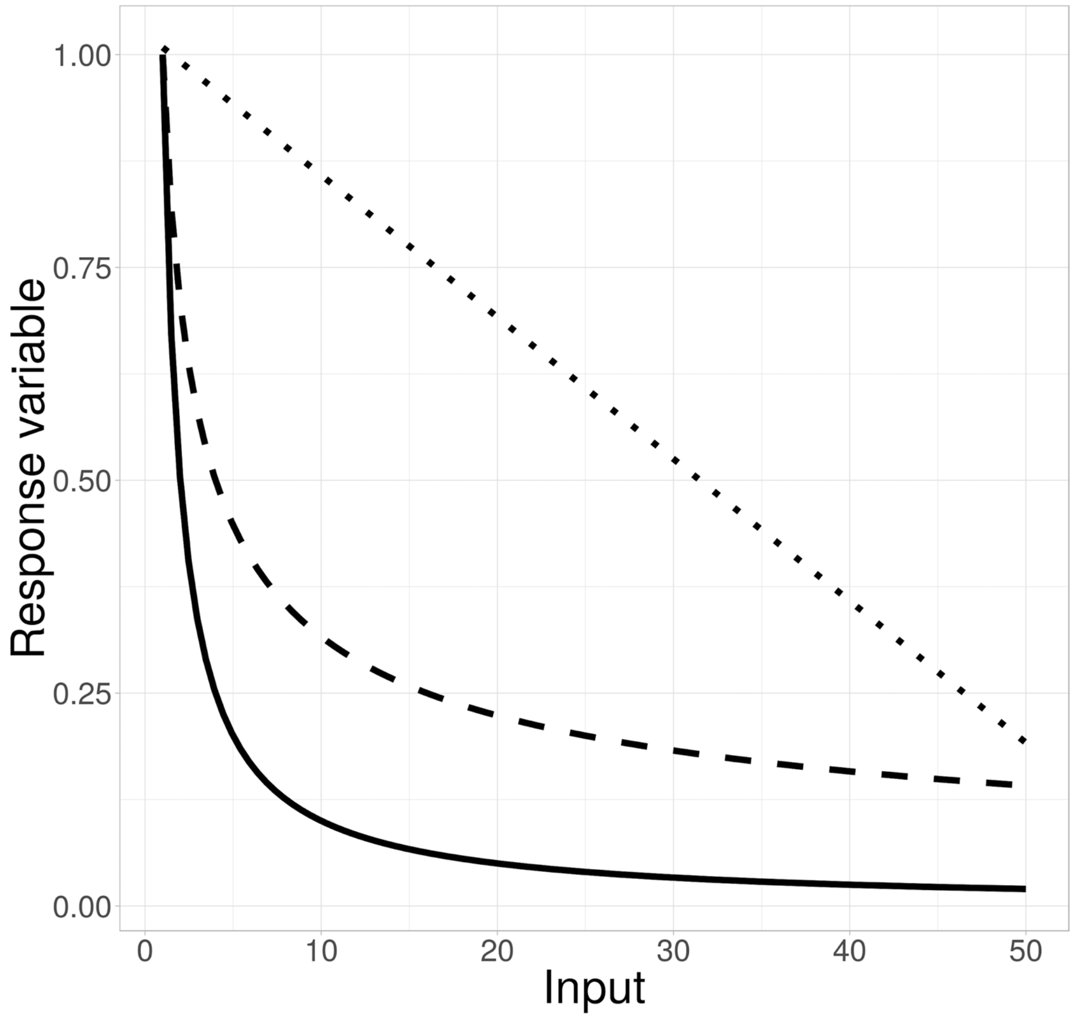


Appendix slides

Quantal regions

A region of a function in which **large variation (error)** in one dimension effects **little response** in some other (task) dimension

- Solid line: strongly quantal
- Dashed line: fairly quantal
- Dotted line: not quantal



Simulation 1 & 2: Muscle sets and ranges

	OOPs	OOPi	OOMs	OOMi	MENT	RIS	LLSAN	LLS
Bilabial	–	–	30	30	20	20	–	–
Labiodental	–	–	–	26	26	26	36	50
Rounded	40	40	–	–	–	–	–	–

Table 1: Maximum muscle stress (kPA) used for the three lip constrictions.

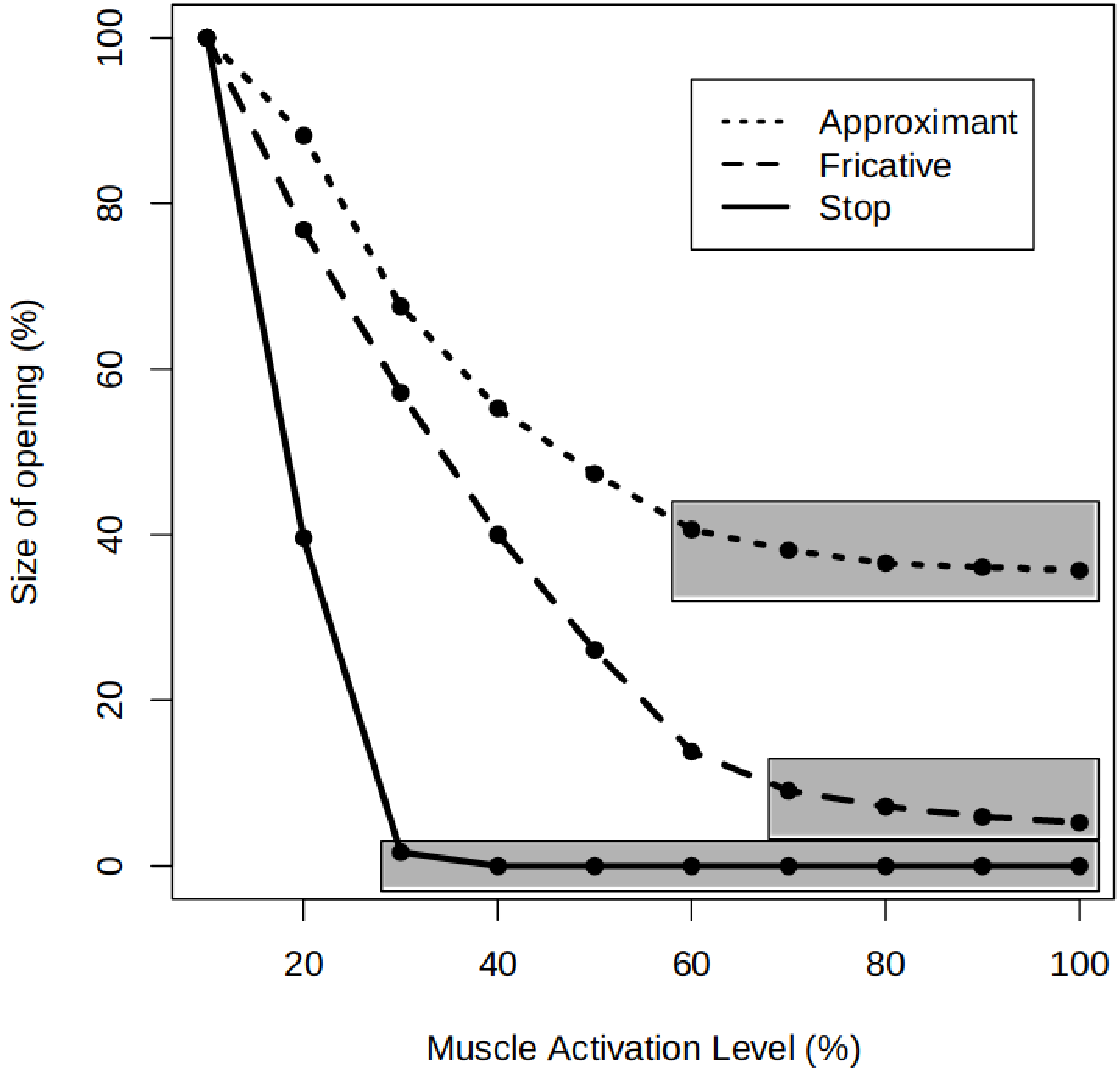
OOMs/i: superior/inferior marginal orbicularis oris
OOPs/i: superior/inferior peripheral orbicularis oris
MENT: mentalis
RIS: risorius
LLSAN: levator labii superioris alaeque nasi
LLS: levator labii superioris

Simulation 2 noise muscles: above muscles, plus depressor anguli oris, buccinator, depressor labii inferior, levator anguli oris, zygomaticus

Simulation 1: Q-scores

The **Q-score** of a function quantifies quantality (Moisik and Gick 2017):

- Compares first derivative in earlier and later ranges
- Based on heuristics in Moisik & Gick (2017):
  - Stop is *strongly quantal*
  - Fricative and approximant are *moderately quantal*



Simulation 1 & 2: Calculating opening size

Simulation 1: Count pixels in coronal images, convert to mm^2

- Labiodental calculated between lower lip and upper teeth
- Other sounds between lower lip and upper lip

Simulation 2: Calculate minimum opening size along a series of cutting planes

- Necessary because of large number of simulations

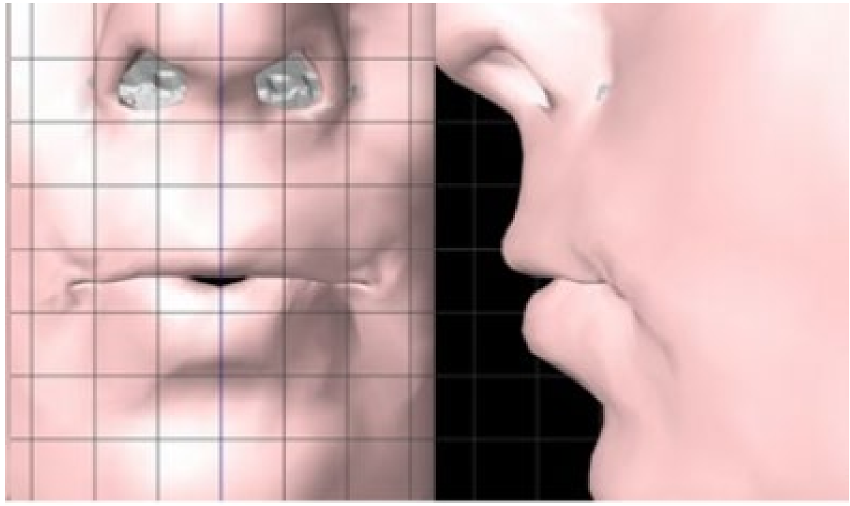
Probabilistic sampling of inputs done using the BatchSim tool

Simulation 2: Robustness to surrounding muscles

Question: Are these postures robust to interference from surrounding muscles?

Focus on approximant (activating OOP)

- No contact, easier to see variable effects



Two types of simulations:

1. Is lip constriction stable when there is surrounding muscle noise?
2. How does degree of OOP activation affect this stability?

Simulation 2: Type 1

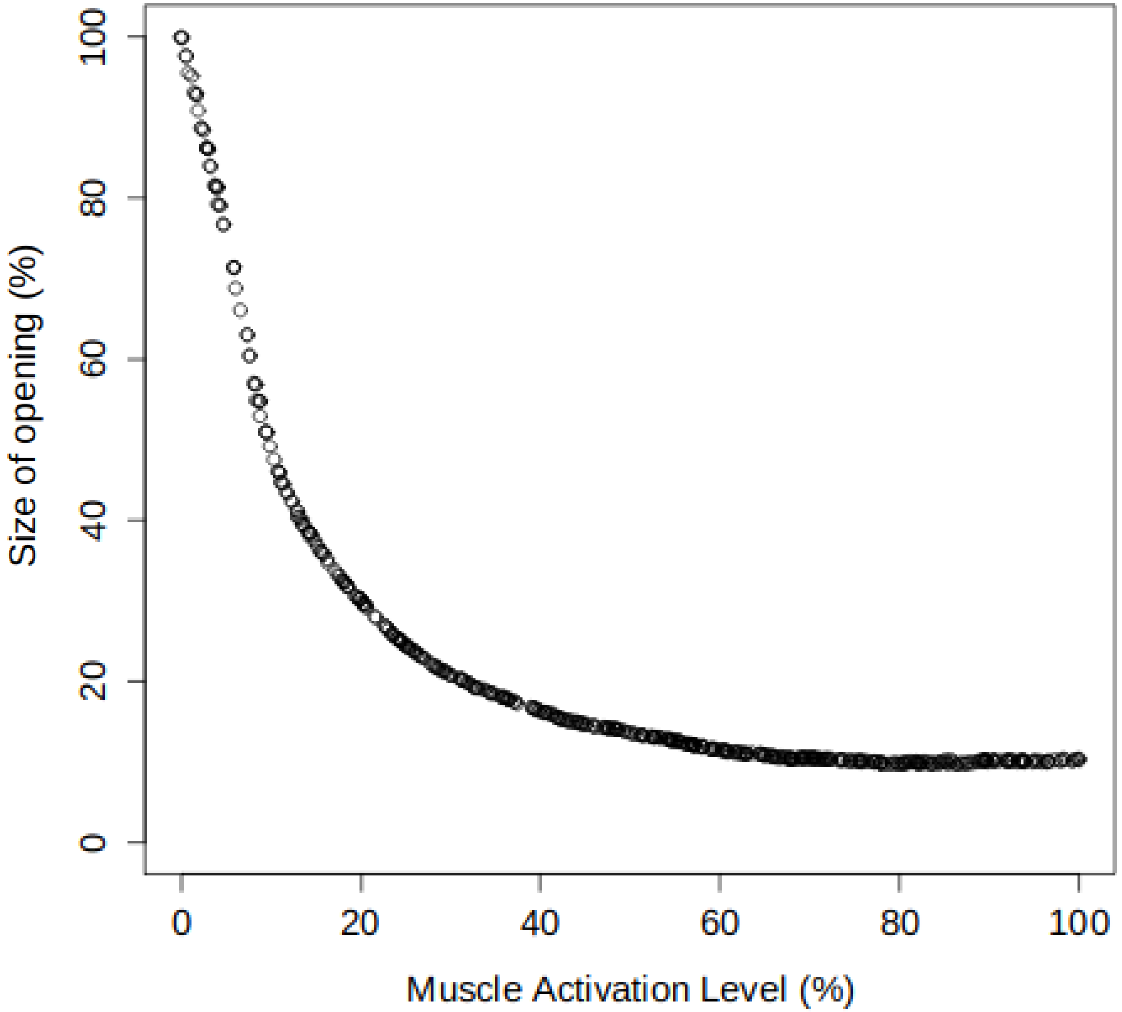
Sampled OOP activation ~ ***U(0%, 100%)***

1. Without activation of surrounding muscles (same as Sim. 1)

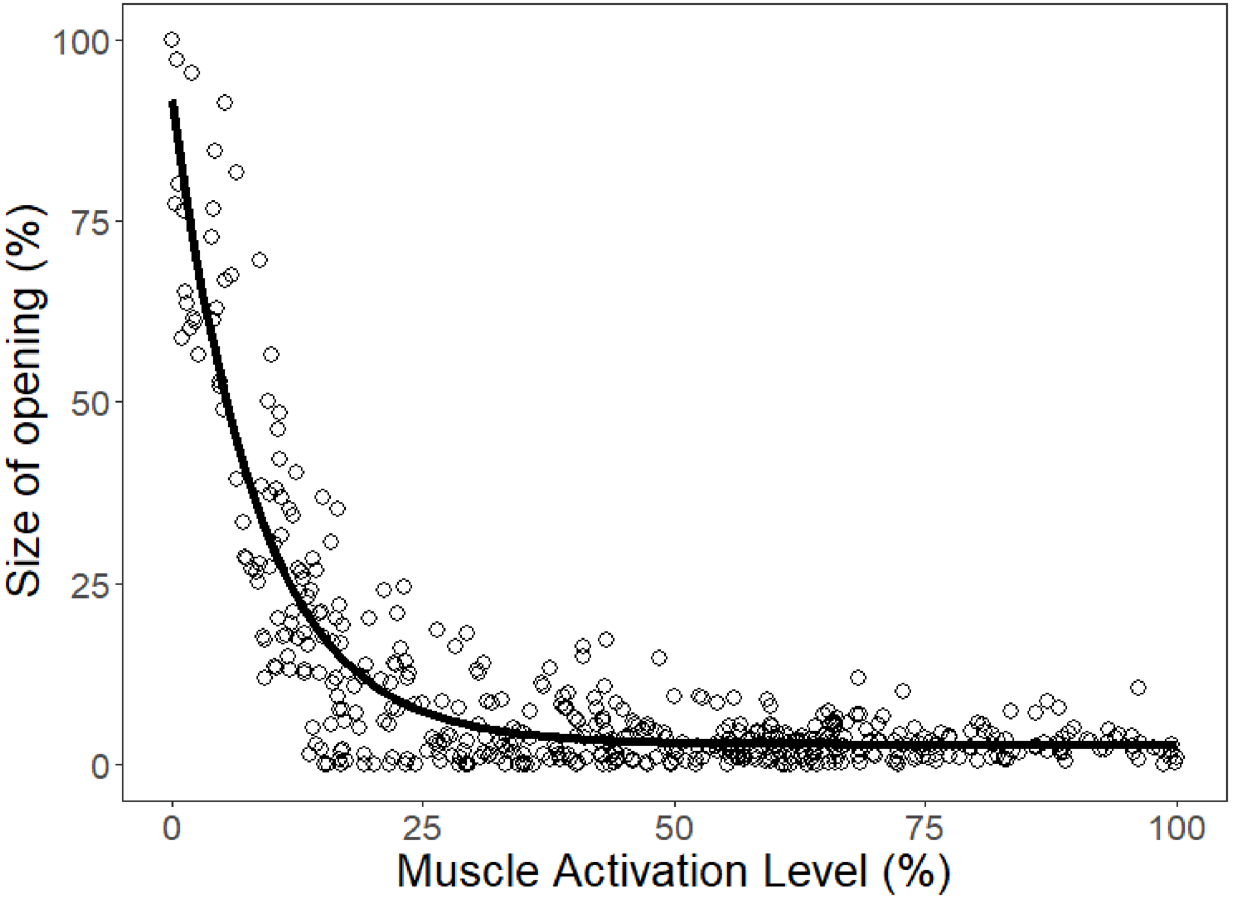
1. With activation of surrounding muscles ~ ***U(0%, 5%)***

Simulation 2: Type 1 Results

No surrounding noise



Surrounding noise



Simulation 2: Type 2

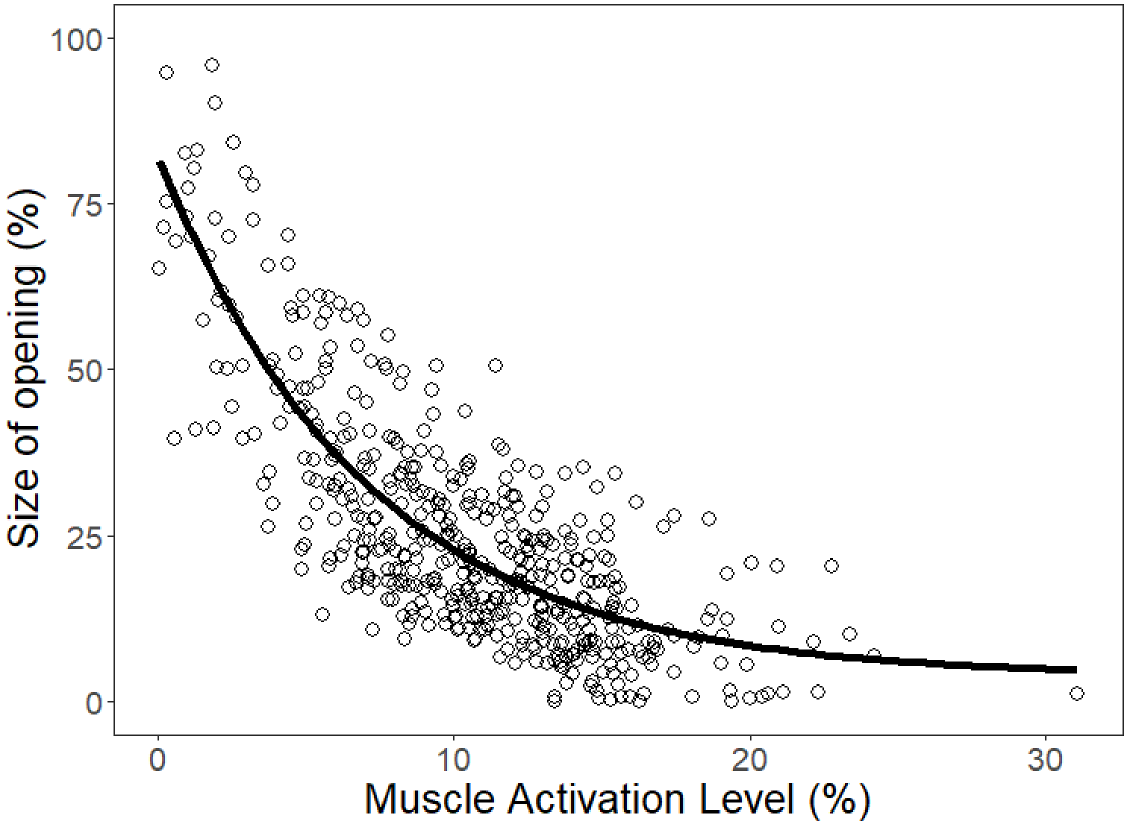
Sampled OOP activation from **two distributions**

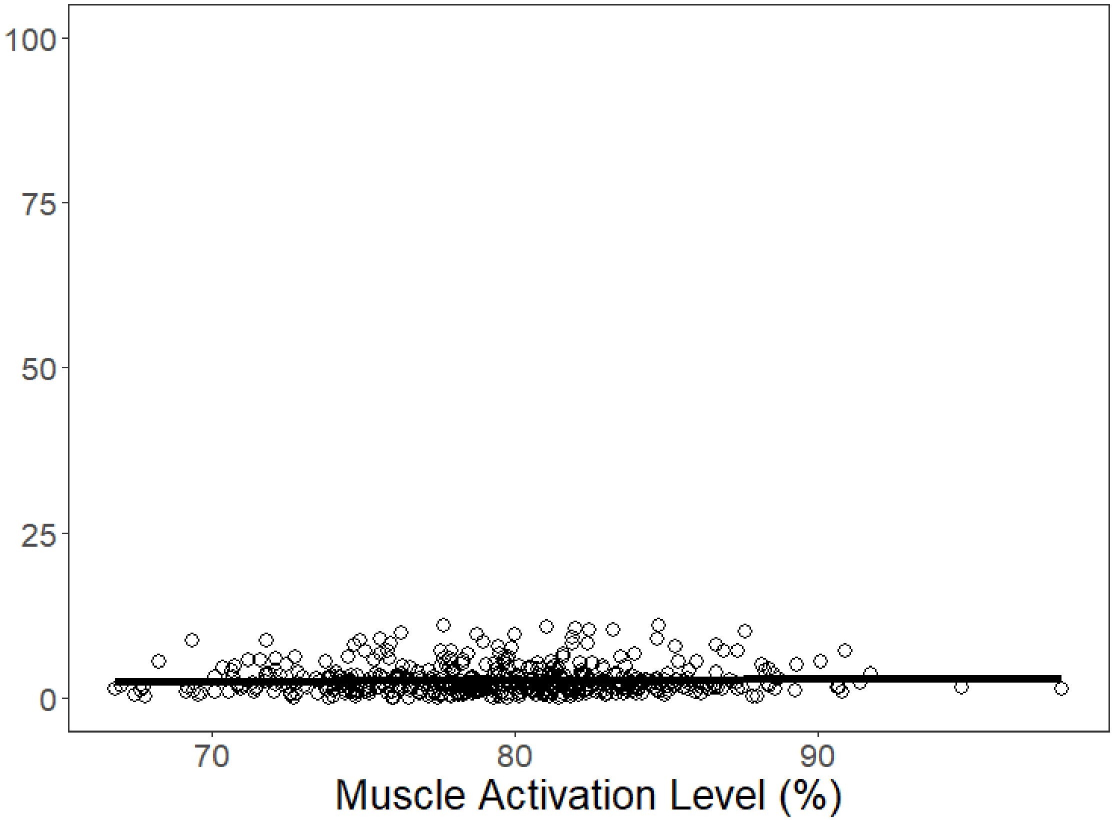
1. **Low activation** $\sim N(\underline{10\%}; 10\%)$

1. **High activation** $\sim N(\underline{80\%}; 10\%)$

Other muscles $\sim U(0\%, 5\%)$

Simulation 2: Type 2 Results





Simulation 2: Type 2 Results

<u>Higher OOP activation reduces interference</u> from surrounding muscles

- Variability in **high activation** region is significantly lower

The high activation region falls in the quantal region in Simulation 1!

- Same region is **robust** to both <u>intrinsic</u> and <u>extrinsic</u> activation noise