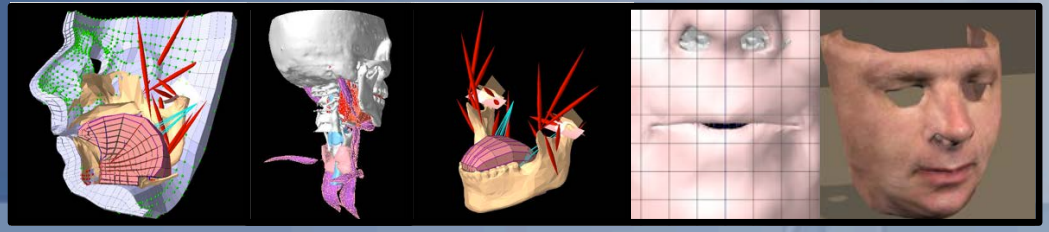


Coarticulation, superposition, representation

Bryan Gick

University of British Columbia & Haskins Labs

With thanks to my collaborators in this work, including: Blake Allen, Peter Anderson, Pierre Badin, Lauretta Cheng, Chenhao Chiu, Donald Derrick, Jonathan deVries, John Esling, Sidney Fels, Cormac Flynn, Naomi Francis, Judith Hall, Megan Keough, Anna Klenin, Ekaterina Komova, Ho Beom Kwon, Yadong Liu, John Lloyd, Sophia Luo, Connor Mayer, Ezra Mizrahi, Scott Moisik, Doug Pulleyblank, Kate Radford, François Roewer-Despres, Murray Schellenberg, Arian Shamei, Pat Shaw, Ian Stavness, Ryan Taylor, Denise Tom, Felicia Tong, Ling Tsou, Eric Vatikiotis-Bateson, Doug Whalen & many others

Labex International Distinguished Chair Seminar Series, Part 3. Paris, France. 28 *March 2019*

Overview of Seminar…

Week 1: The Parts of Speech I: Dimensionality and Modularization

- READINGS: DRAFT Ch 1 of *Embodying Speech: How Bodies Talk*
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators
  - Cats/hammers/frogs & functional body parts/devices
  - The transformation of "coordinative structures"

Week 2: The Parts of Speech II: Quantality and Speech Movements

- READINGS: Lip papers, palate paper
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators

Week 3: Coarticulation, Superposition, Representation

- READINGS: coarticulation paper
  - Devices over time – 4 temporal properties;
  - taps/flaps > motor abundance
  - Coarticulation as overlap/superposition of modules; smile

Embodied Speech workshop

  - Towards units of embodied speech

Week 4: Emergence, Sound Change, Ontogeny and Phylogeny

- READINGS: TBA

Overview of Seminar…

Week 1: The Parts of Speech I: Dimensionality and Modularization

- READINGS: DRAFT Ch 1 of *Embodying Speech: How Bodies Talk*
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators
  - Cats/hammers/frogs & functional body parts/devices
  - The transformation of "coordinative structures"

Week 2: The Parts of Speech II: Quantality and Speech Movements

- READINGS: Lip papers, palate paper
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators

Week 3: Coarticulation, Superposition, Representation

- READINGS: coarticulation paper
  - Devices over time – 4 temporal properties;
  - taps/flaps > motor abundance
  - Coarticulation as overlap/superposition of modules; smile

Embodied Speech workshop

  - Towards units of embodied speech

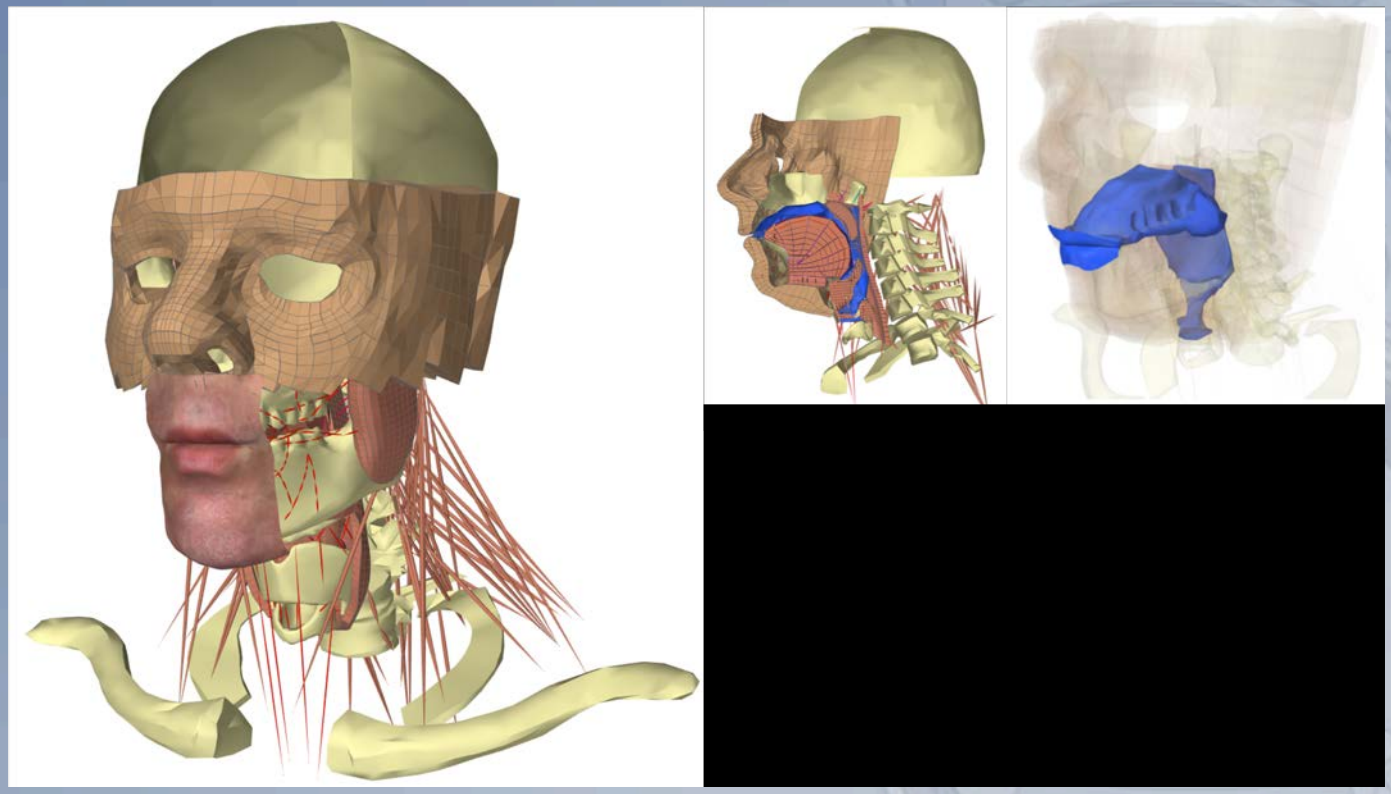
Week 4: Emergence, Sound Change, Ontogeny and Phylogeny

- READINGS: TBA

Frank Model in ArtiSynth

We created Frank to understand dimensionality of the vocal tract

State-of-the-art platform for biomechanical head/face/VT modeling



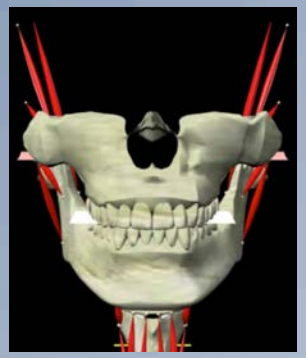
artisynth.org

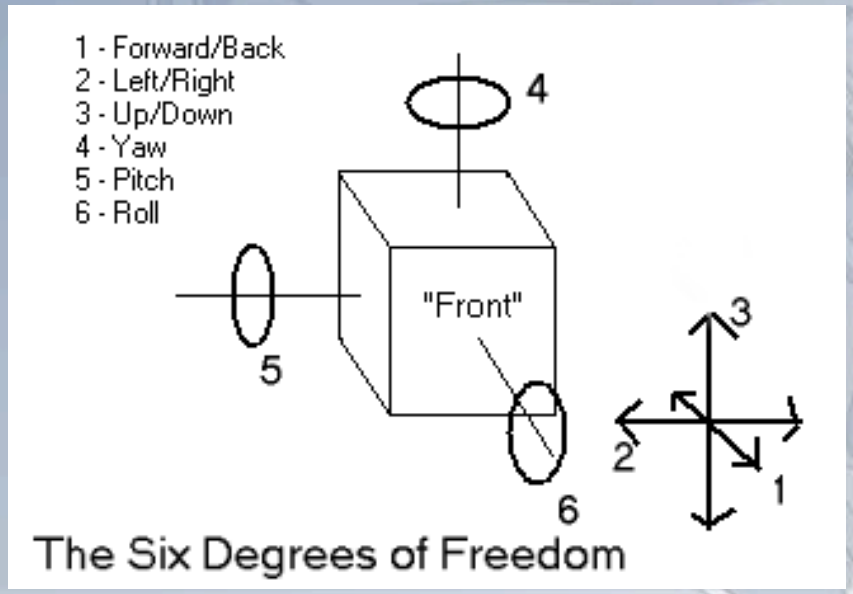
Dimensionality Problem

The human body has practically unbounded degrees of freedom

Mechanical **degrees of freedom** (DOF) = independent parameters that define a mechanical system's possible movements

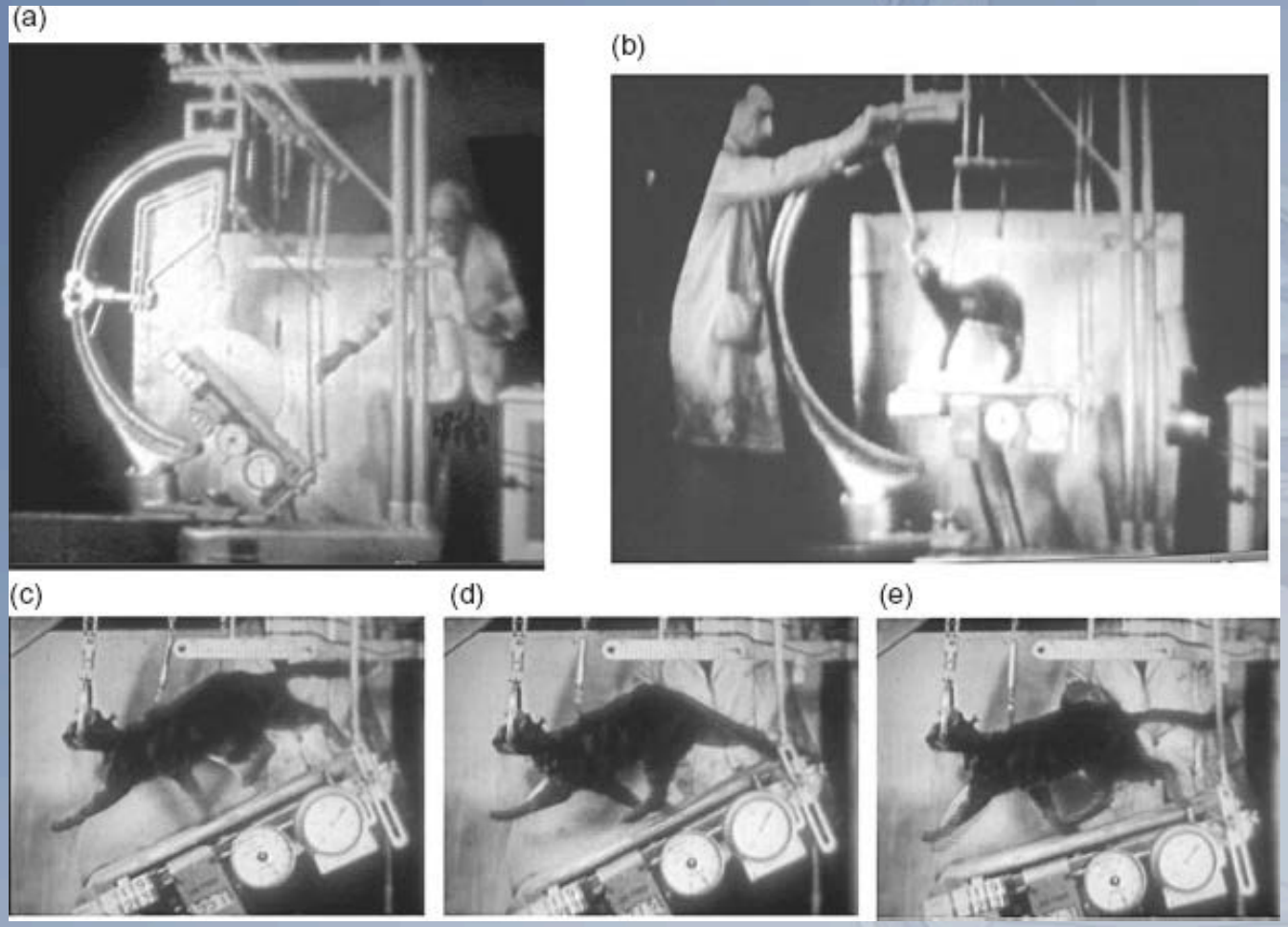
How many DOFs does a "rigid body" like the jaw have?





The Six Degrees of Freedom

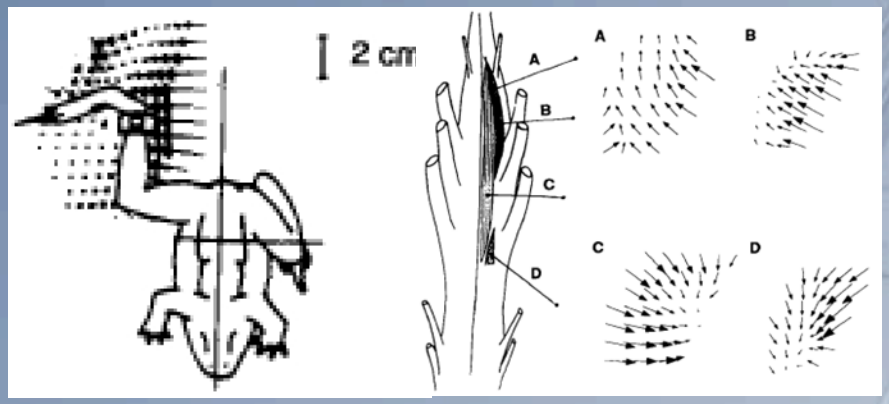
Where do complex actions live?



Sherrington & Graham Browns' cats (1910-1930's)

Neurophysiological Modules

We've long known of functional modules in motor systems

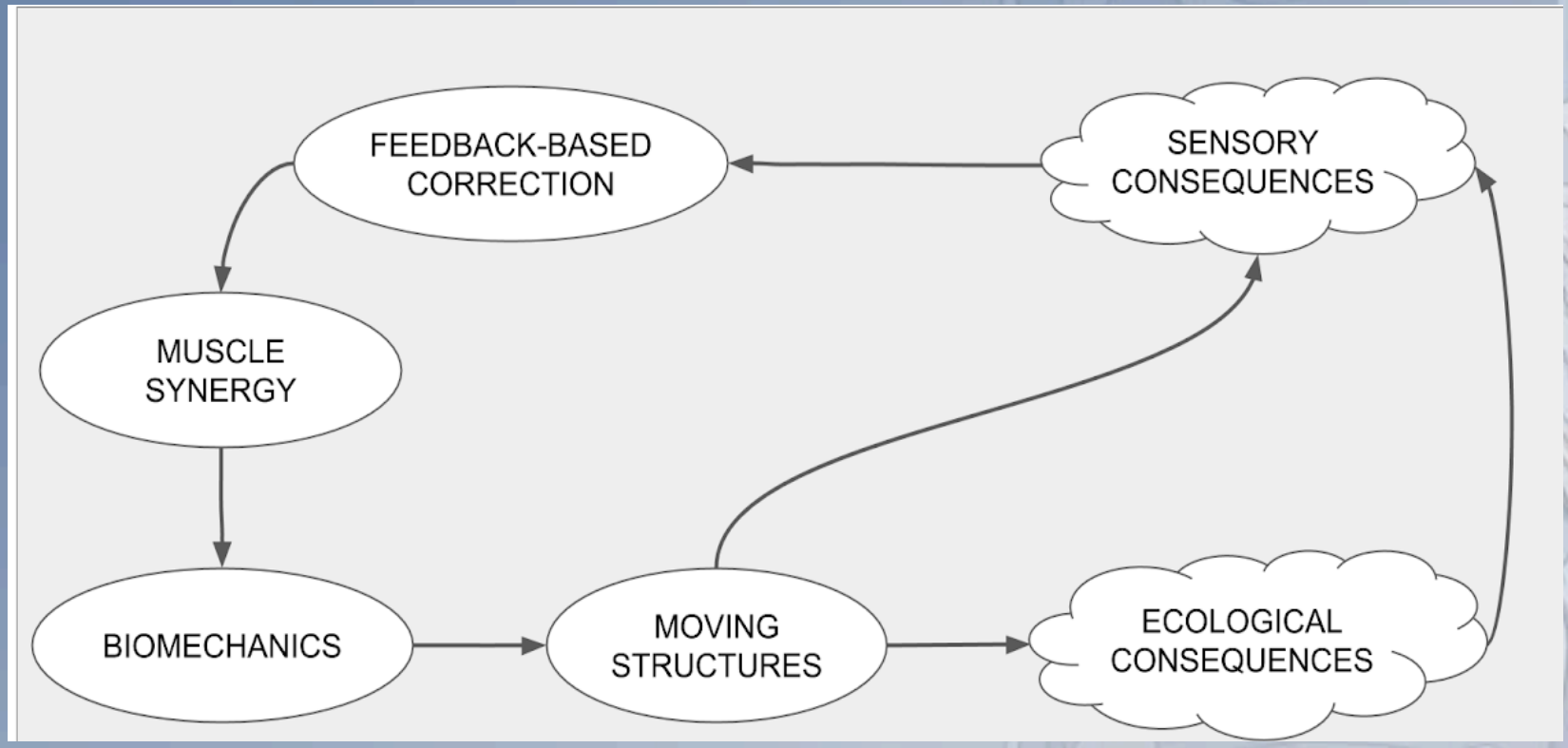


"spinalized" frogs
Bizzi et al. (*Science* 1991)

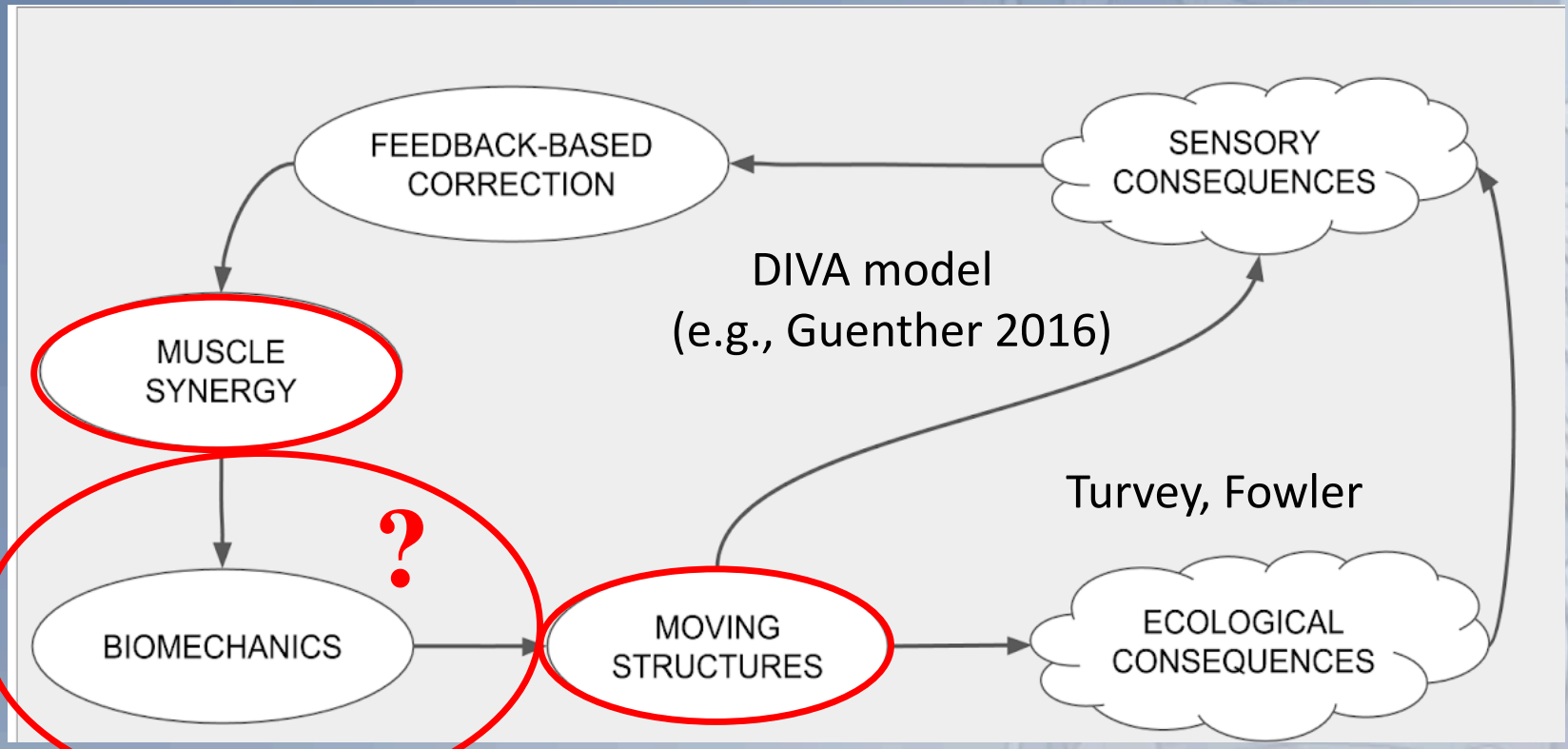
- Solution to Berstein's dimensionality ("DOF") problem
- Neural pathways corresponding to *useful* and *reliable* actions

…so, what are some useful and reliable actions in speech?

“Whole“ modules/devices



“Whole“ modules/devices



Overview of Seminar...

Week 1: The Parts of Speech I: Dimensionality and Modularization

- READINGS: DRAFT Ch 1 of *Embodying Speech: How Bodies Talk*
- TOPIC: Embodied units of representation in embodied speech: Why we need them and how they work
  - brief intro to ArtiSynth & the dimensionality problem
  - midsagittal/articulator representations in phonetics
  - Sherrington/Graham Brown cats/Bernstein hammers/Bizzi frogs
  - functional body parts/devices
  - the transformation of "coordinative structures"

Week 2: The Parts of Speech II: Quantality and Speech Movements

- READINGS: Lip papers, palate paper

Week 3: Coarticulation, Superposition, Representation

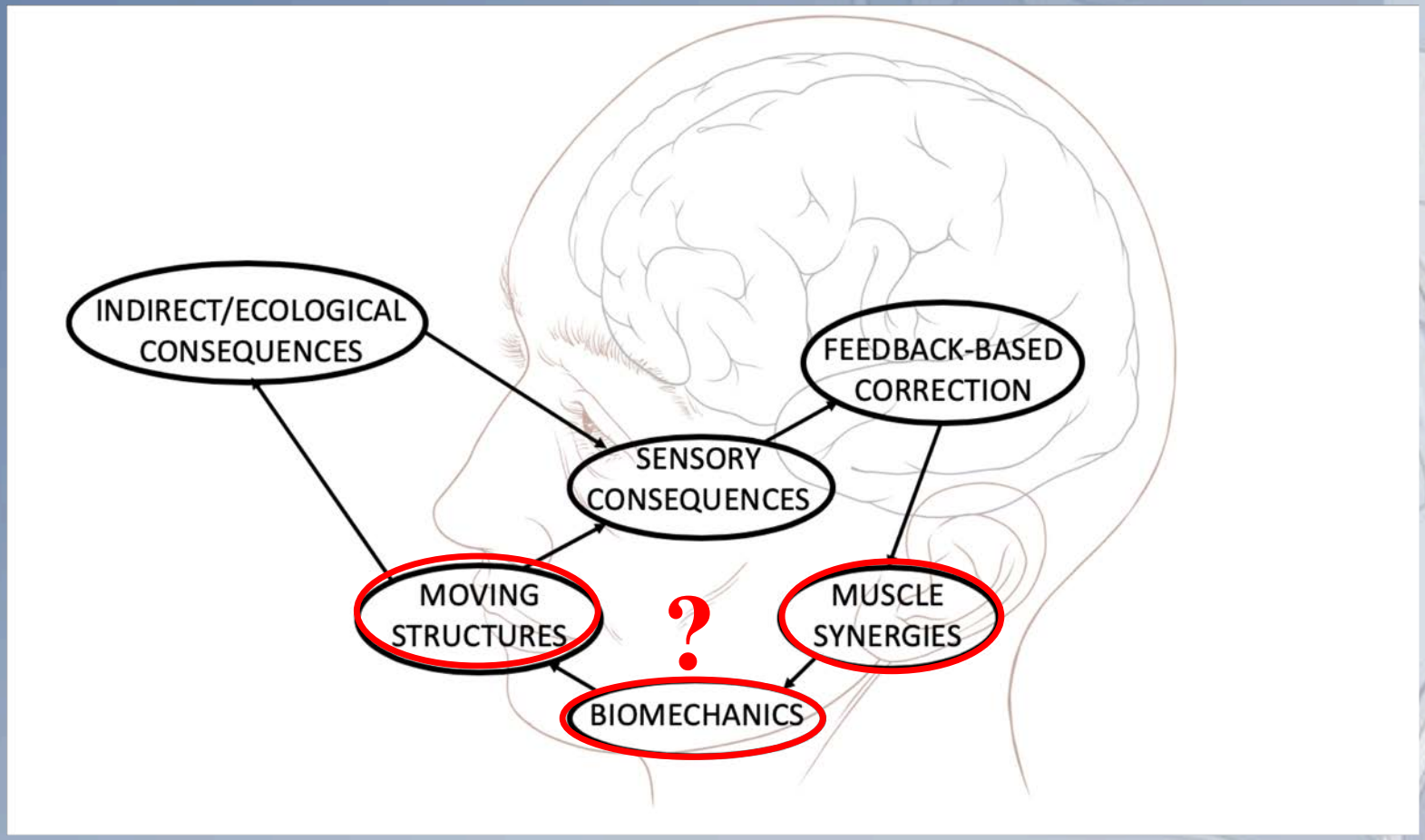
- READINGS: DRAFT Ch 5, etc. TBA on coarticulation

Embodied Speech workshop

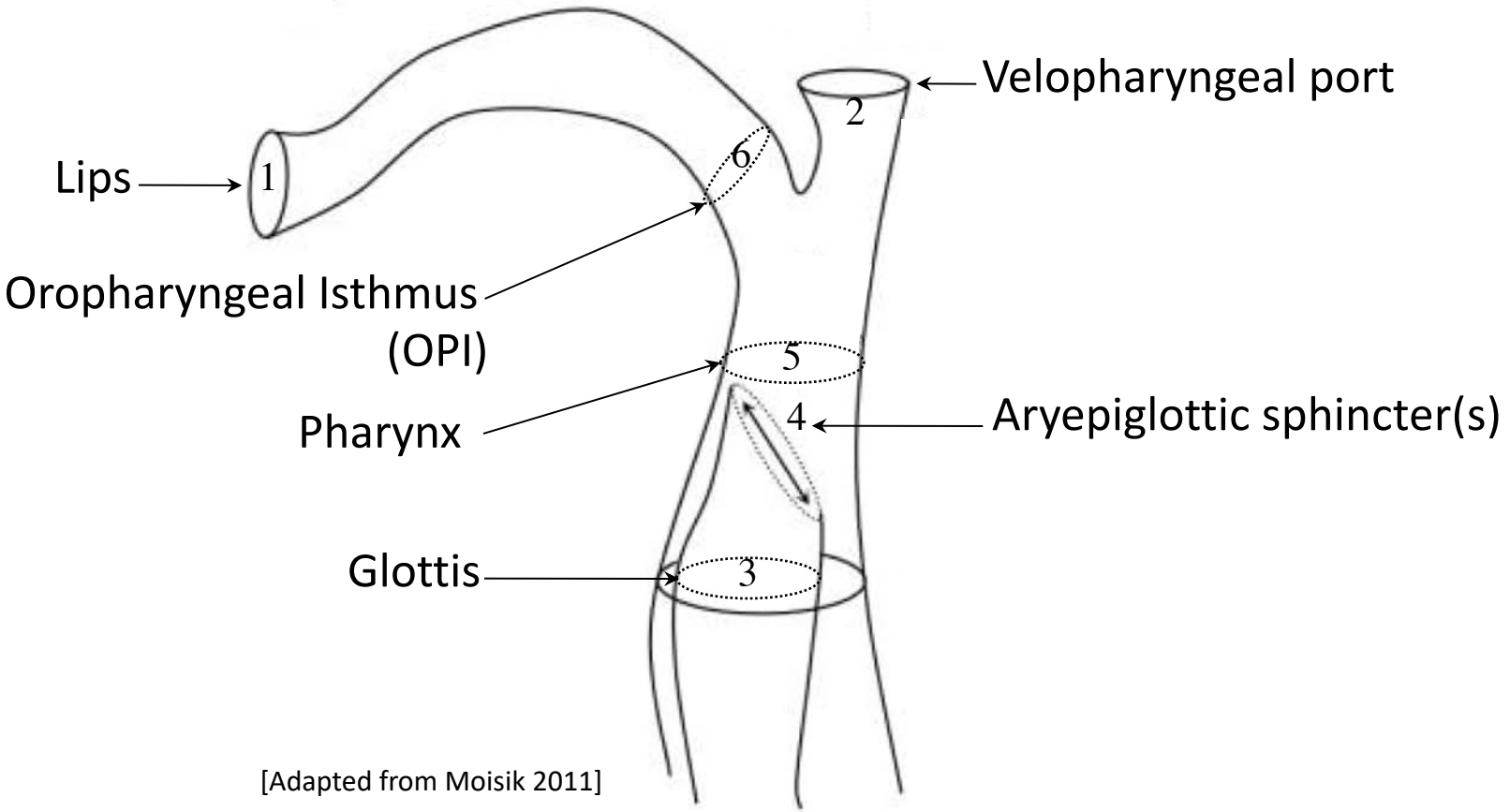
Week 4: Emergence, Sound Change, Ontogeny and Phylogeny

- READINGS: TBA

“Whole“ modules/devices



Locations of some transient devices

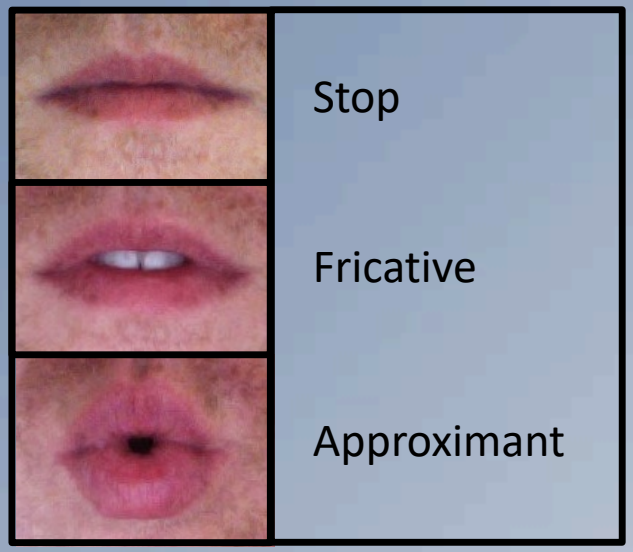


[Adapted from Moisik 2011]

Lips

Lip constrictions for speech movements are *specialized postures*

- each produces a specific-sized opening



Quantal biomechanics

Fujimura (1989: 89) on Stevens' (1989) Quantal Theory: "To create a labial closure, for example, it is not necessary to achieve precise approximation of the lips so that they just touch one another; rather, all one has to do is to make the articulatory force sufficiently great to obtain any complete closure. Likewise in the case of labiodental fricatives there is no need for great precision in the force of the contact; the physical characteristics of the lips and teeth guarantee the desired result though the closing gestures vary widely."

Quantal

> categorical, nonlinear

Allow:

(1) variable muscle activations

(2) feed-forward control

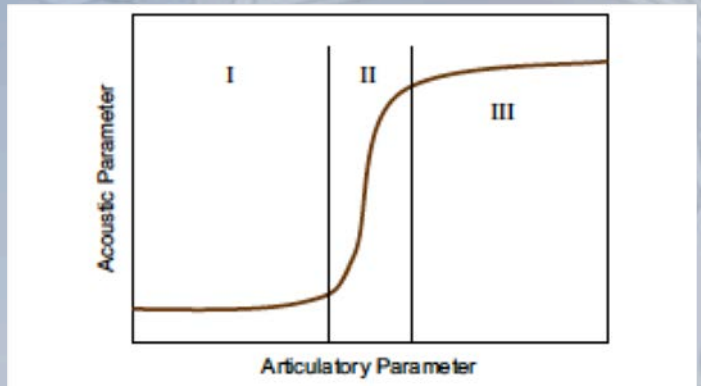
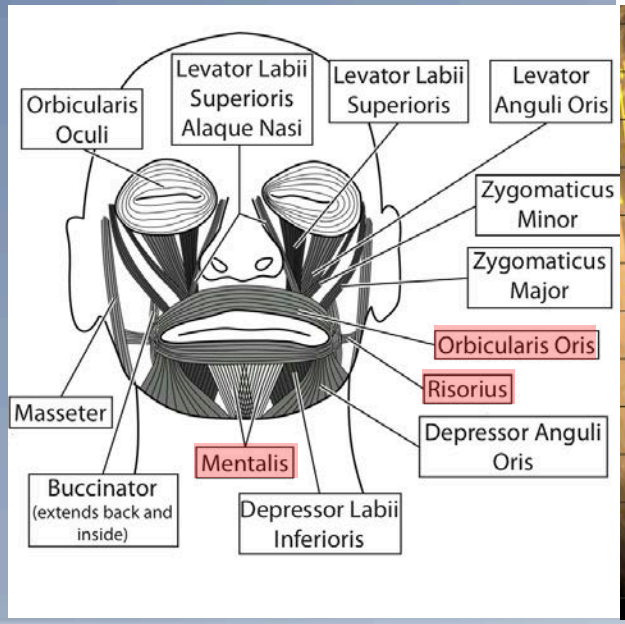


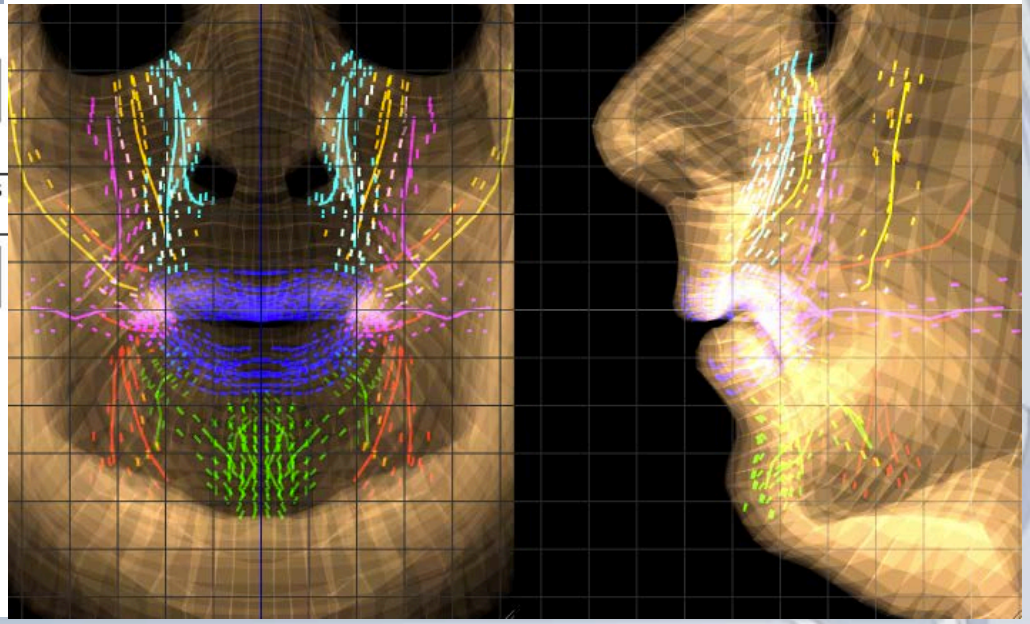
Fig. 1. Hypothetical articulatory/acoustic relation showing two relatively stable regions (I and III) and a region where there is a rapid change in an acoustic parameter for a relatively small change in the articulatory parameter.

(fr. Stevens & Keyser 2010)

Simulation

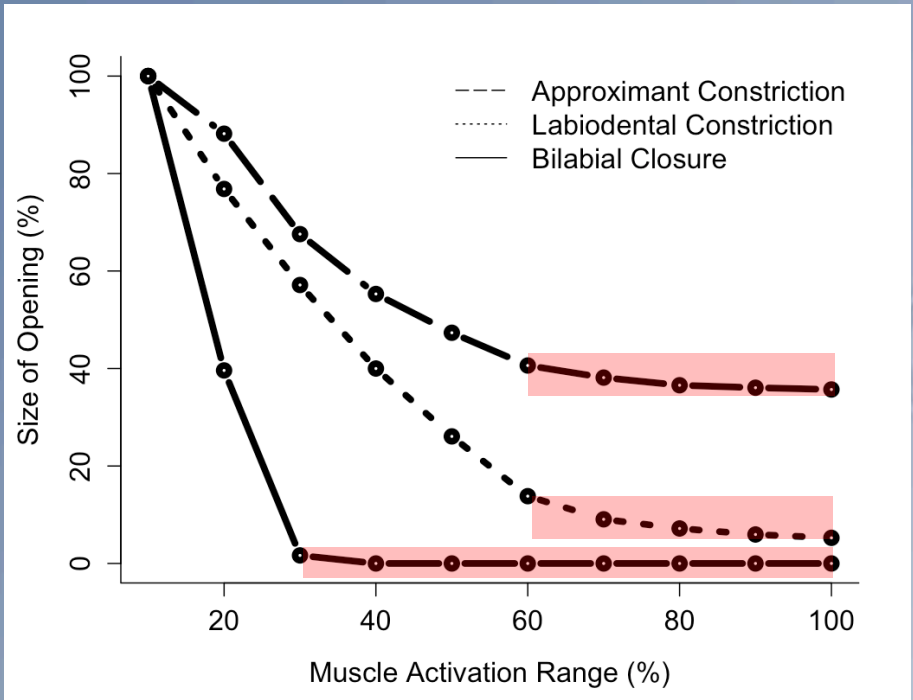
	Bilabial Closure	Labiodental	Approximant
OOPs	--	--	40
OOPi	--	--	40
OOMs	30	--	--
OOMi	30	26	--
MENT	20	26	--
RIS	20	26	--
LLSAN	--	36	--
LLS	--	50	--

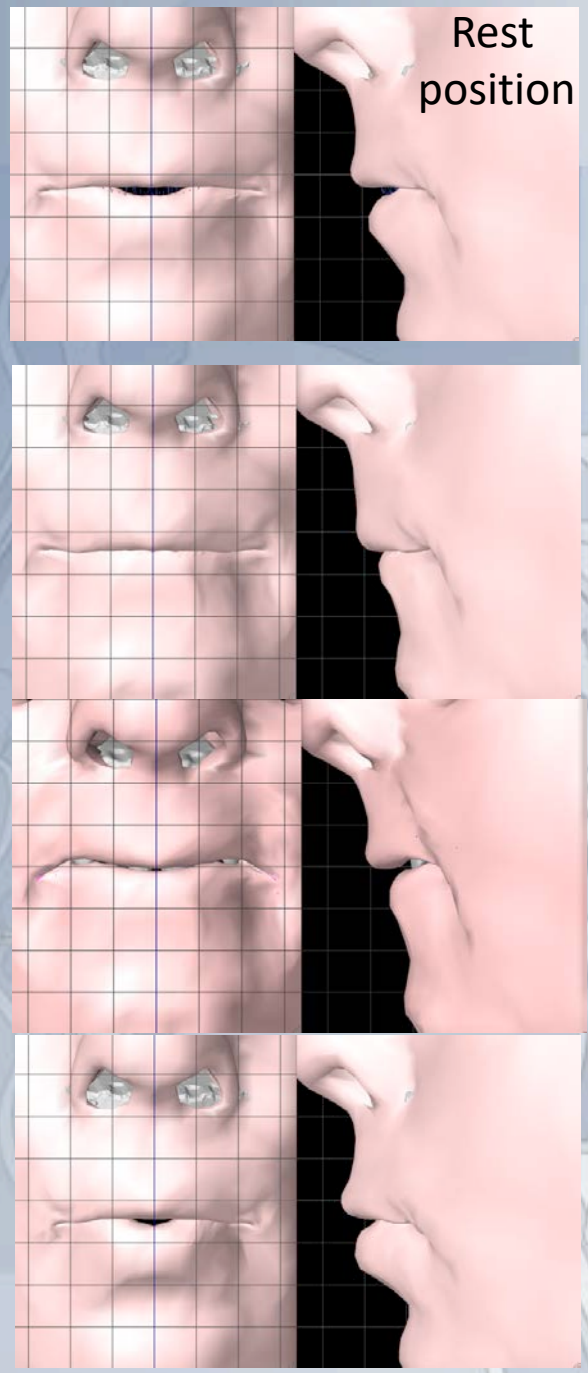




Simulation

Adapted from
Gick & al. (2011) *Canadian Acoustics*





Stable "Quantal" regions shaded in pink

- stop = ballistic overshoot
- fricative = same as stop, but irregular surface
- approximant = saturation

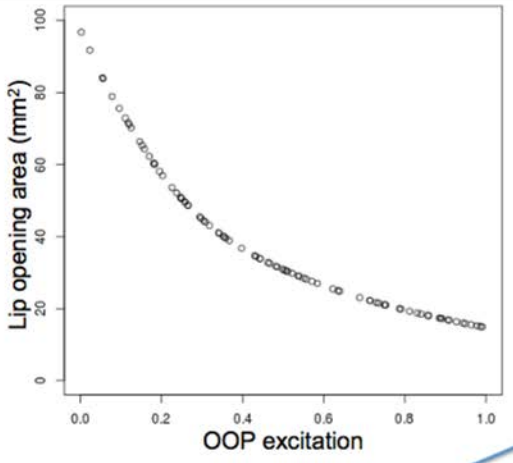
These structures are robust to activation noise…

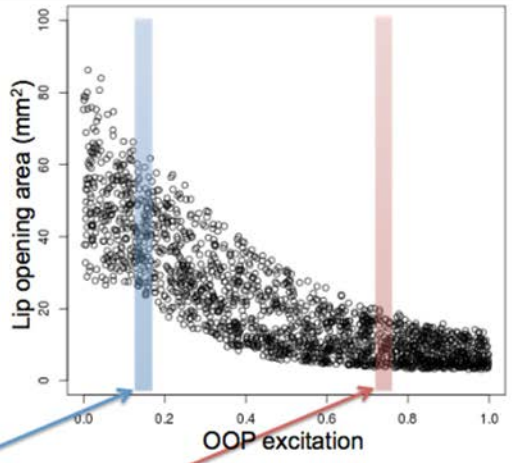
Simulation

Also robust to perturbation from surrounding activations…

Simulations with uniform distribution of OOP excitation, without (left) and with (right) excitation of other face muscles.

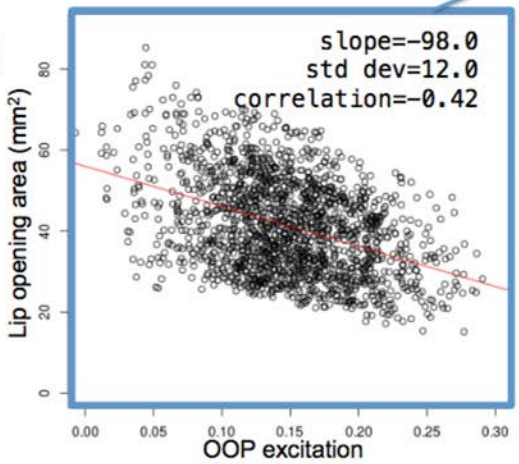
```
(both)  OOP=Uniform(0,1.0)
(left)  other=0.0
(right) other=Uniform(0,0.5)
```

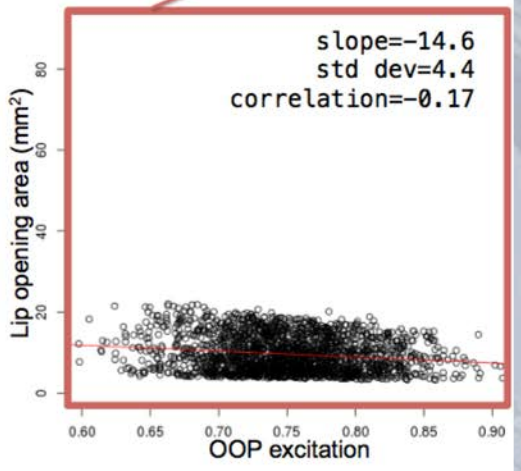




Simulations with low (left) and high (right) OO excitation along with excitation of other face muscles.

```
(left)  OOP=Normal(0.15, 0.1)
(right) OOP=Normal(0.75, 0.1)
(both)  other=Uniform(0,0.5)
```



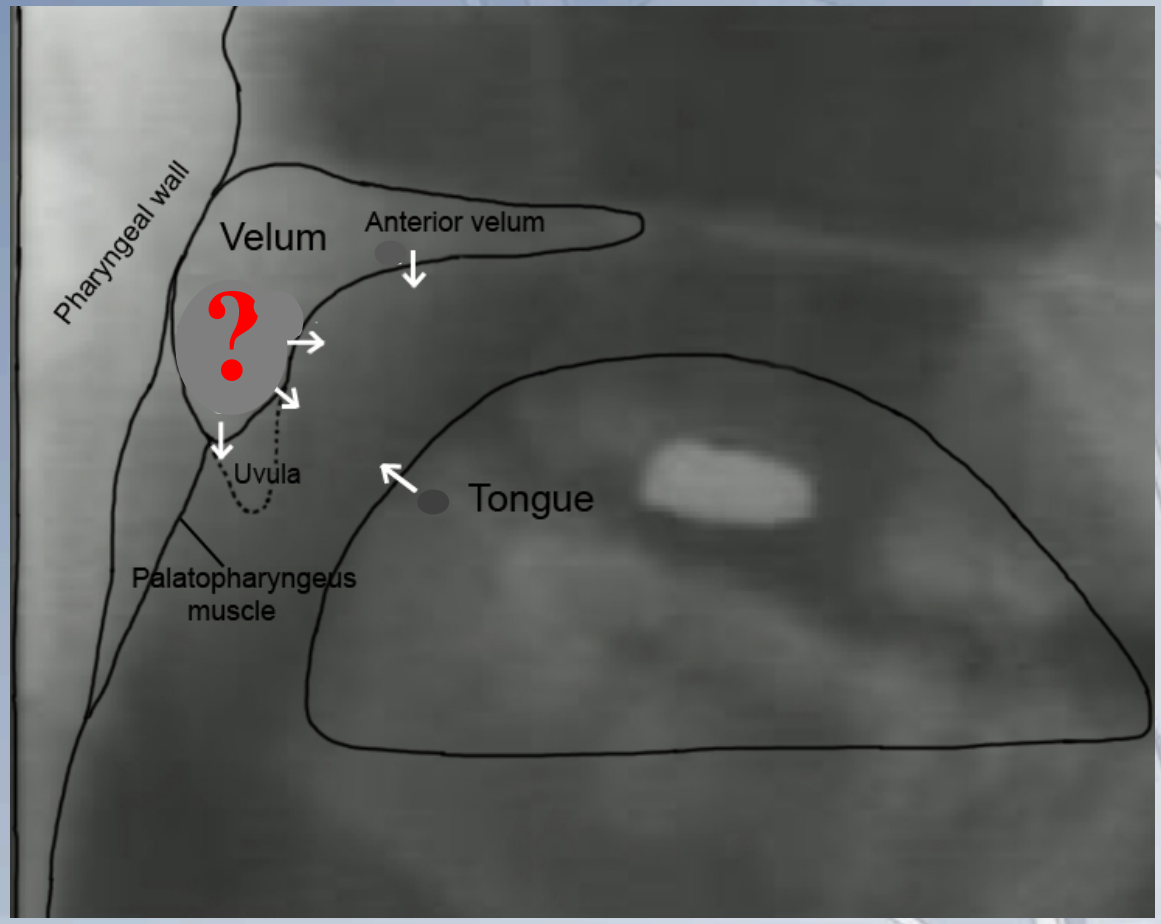


Stavness I, Roewer-Despres F, Gick B. (*JASA* 2016)

French /ʁ/ Experiment

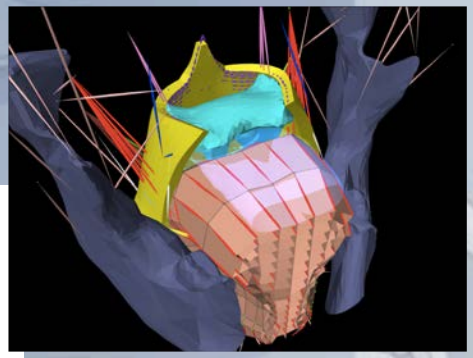
QUESTION:

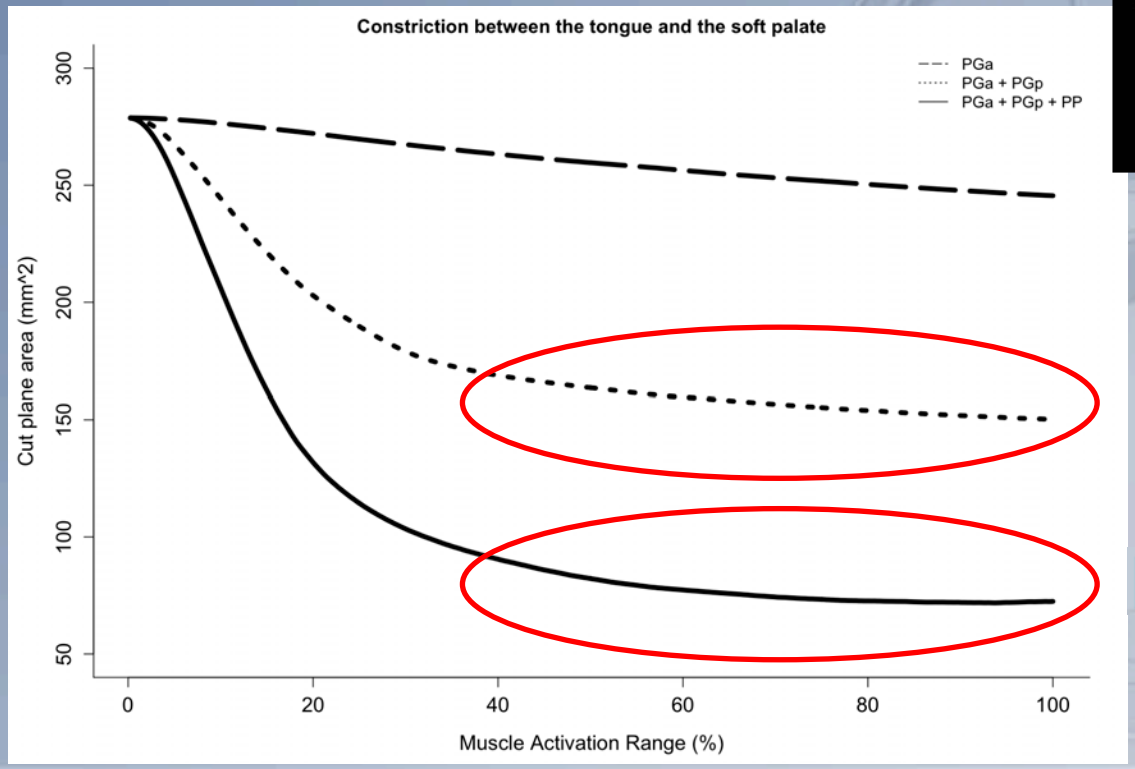
- What is this?



Similar structures in the palate

More simulation results
– not all modules are created equal…





[u]

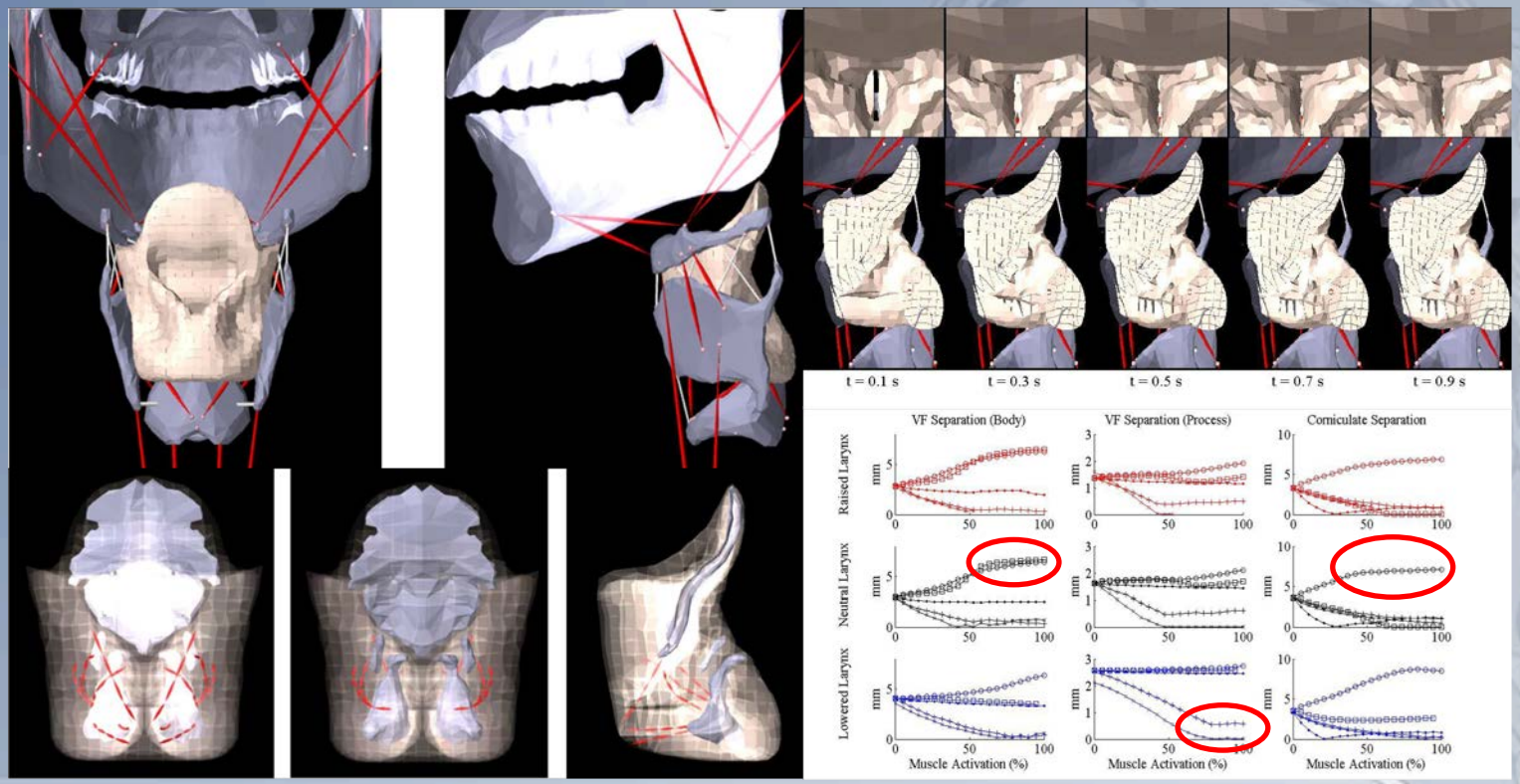
[ʁ]

Gick & al. (2014)

...and the Larynx

Same principles apply to phonation types/laryngeal states…

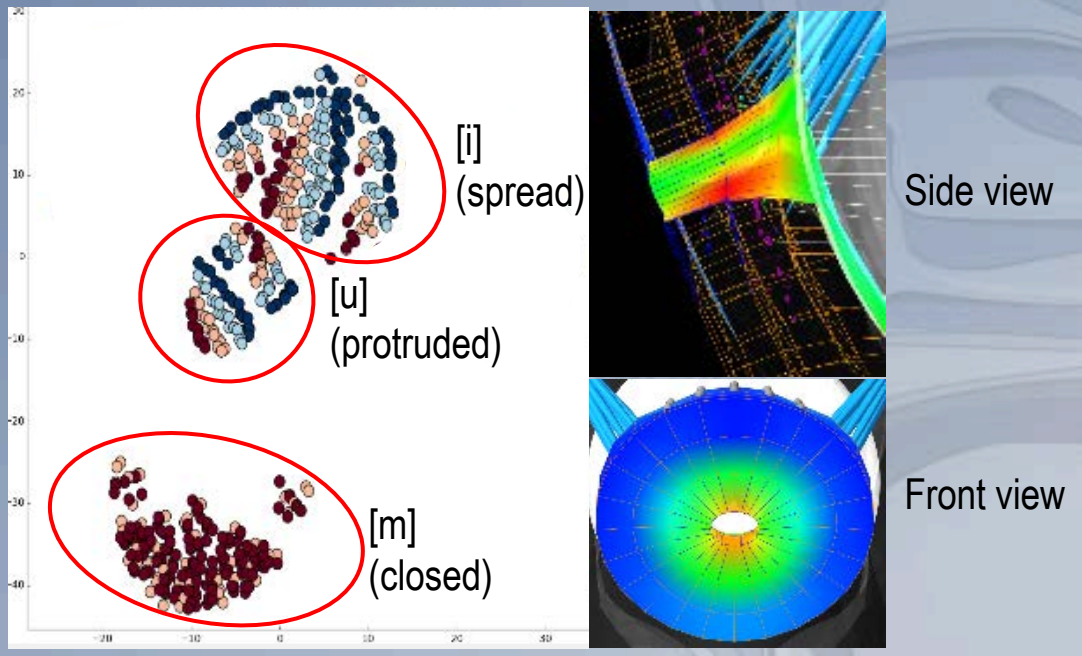
- Moisik & Gick (2017)



All of these structures are robust to activation noise!

Emergent primitives in imitation:

...also robust to "spatial" noise



t-SNE visualizations of a portion of the (symmetrical) biomechanical space

Each data point corresponds to an individual simulation

...this is just what (babies') mouths do!

Overview of Seminar…

Week 1: The Parts of Speech I: Dimensionality and Modularization

- READINGS: DRAFT Ch 1 of *Embodying Speech: How Bodies Talk*
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators
  - Cats/hammers/frogs & functional body parts/devices
  - The transformation of "coordinative structures"

Week 2: The Parts of Speech II: Quantality and Speech Movements

- READINGS: Lip papers, palate paper
  - Embodied units of representation in embodied speech
  - ArtiSynth & the dimensionality problem; midsagittal/articulators

Week 3: Coarticulation, Superposition, Representation

- READINGS: coarticulation paper; taps/flaps paper
  - Devices over time – 4 temporal properties;
  - taps/flaps > motor abundance
  - Coarticulation as overlap/superposition of modules; smile

Embodied Speech workshop

  - Towards units of embodied speech

Week 4: Emergence, Sound Change, Ontogeny and Phylogeny

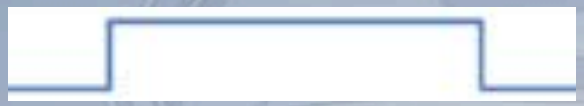
- READINGS: TBA

Devices over time

4 temporal properties of devices:

- Transient

- Tonic



- Cyclic

- Sequential

*NB: these are schematic kinematic diagrams, not activations!

Adapted from. : Bavandpour et al. (2015). https://www.frontiersin.org/articles/10.3389/fnins.201500409/full

Transient Devices

- Most devices we've talked about so far
- Single, minimal events (e.g., a constriction)
- Short-term, "one-off" consequences
- *limited, fixed duration*

Tonic Devices

- Similar to transient devices, but…
- *Maintain a state* over a longer duration
- E.g.: facial expression, articulatory setting, tongue bracing, states of the glottis, etc.

Tonic Device: Tongue Bracing

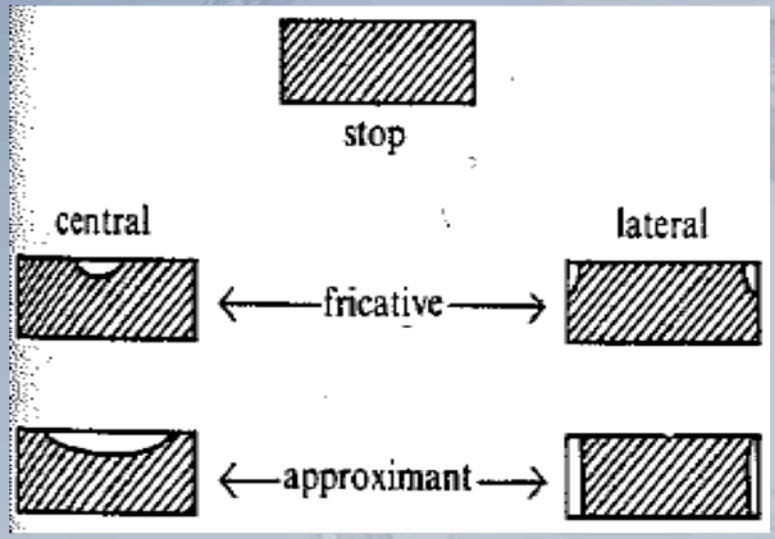
Consider how the tongue works in speech…

The lateral tongue constricts against the hard palate and teeth

…this forms the "aeroacoustic tube"

…also crucial for mechanical stabilization (Brunner & al 2005)

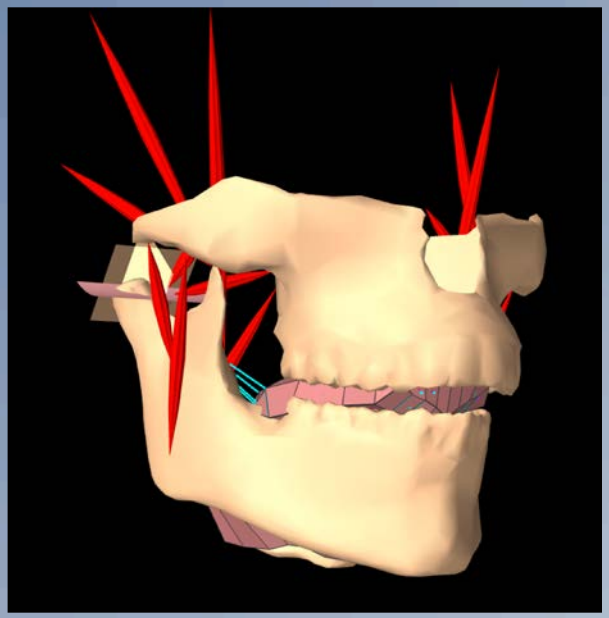
...and for somatosensory feedback about tongue position

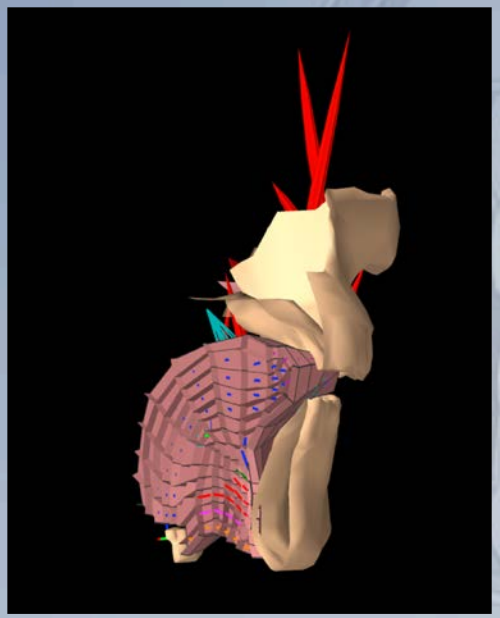


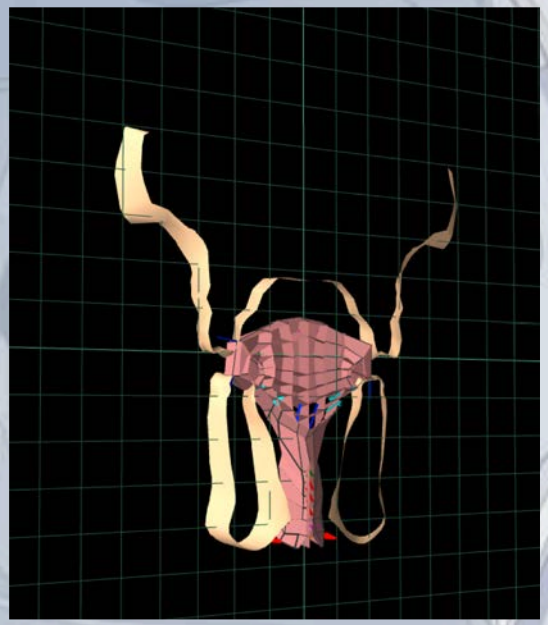
From Catford (1977)

Tongue Bracing

Stone (1990) found that the tongue was braced (against the hard palate and molars) during many lingual constrictions
= “Proxy bone support” like the lips



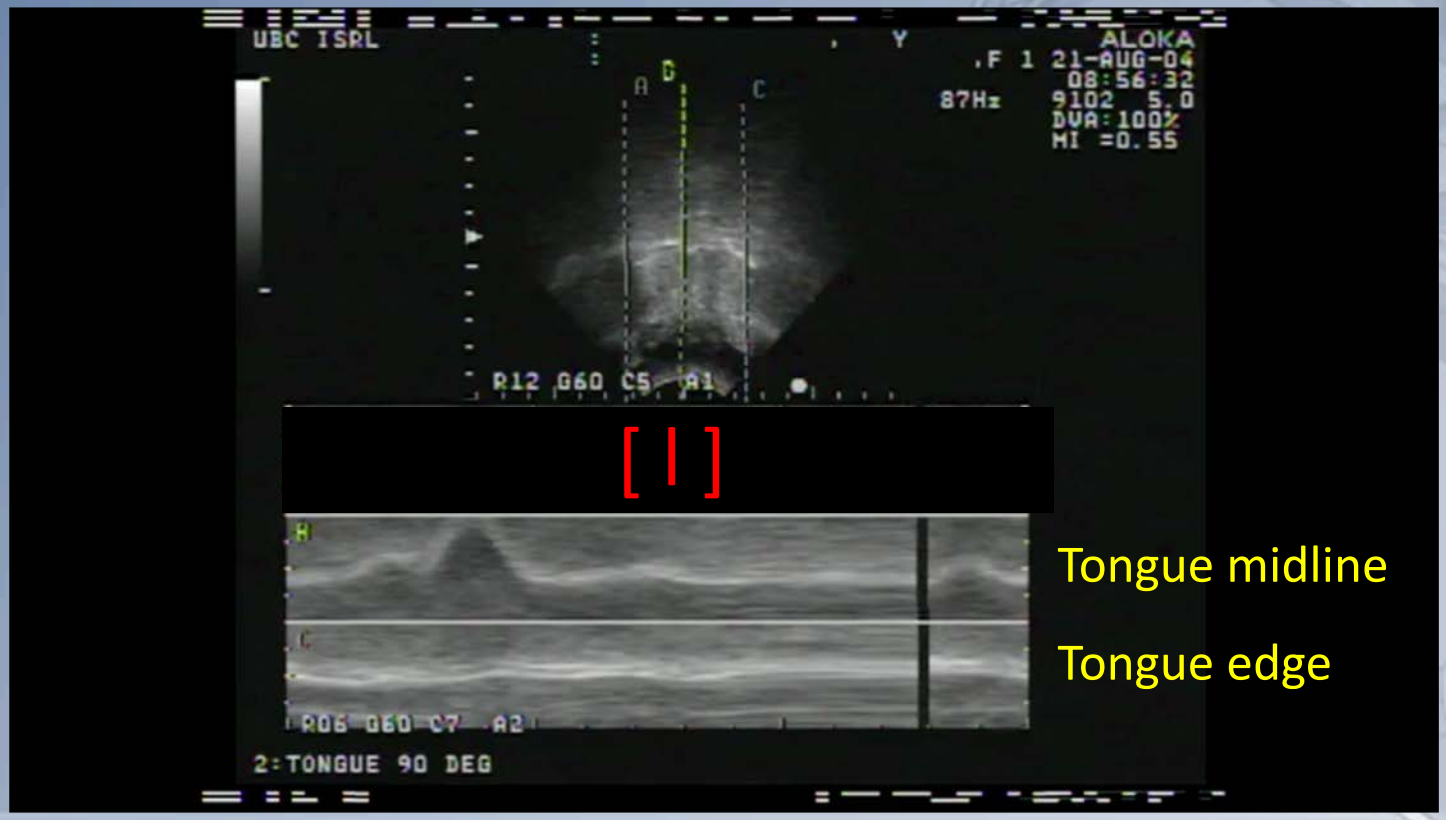




…but how often is it braced? Constant “basis” for speech

The Tongue: Bracing

Here's how bracing works:



Why "bracing" and not just "contact"?

Cyclic Devices

- Similar to tonic devices, but…
- Continued activation produces *periodic output*
- E.g.: respiration, jaw cycle, tongue flapping, sign language arm movements, etc.

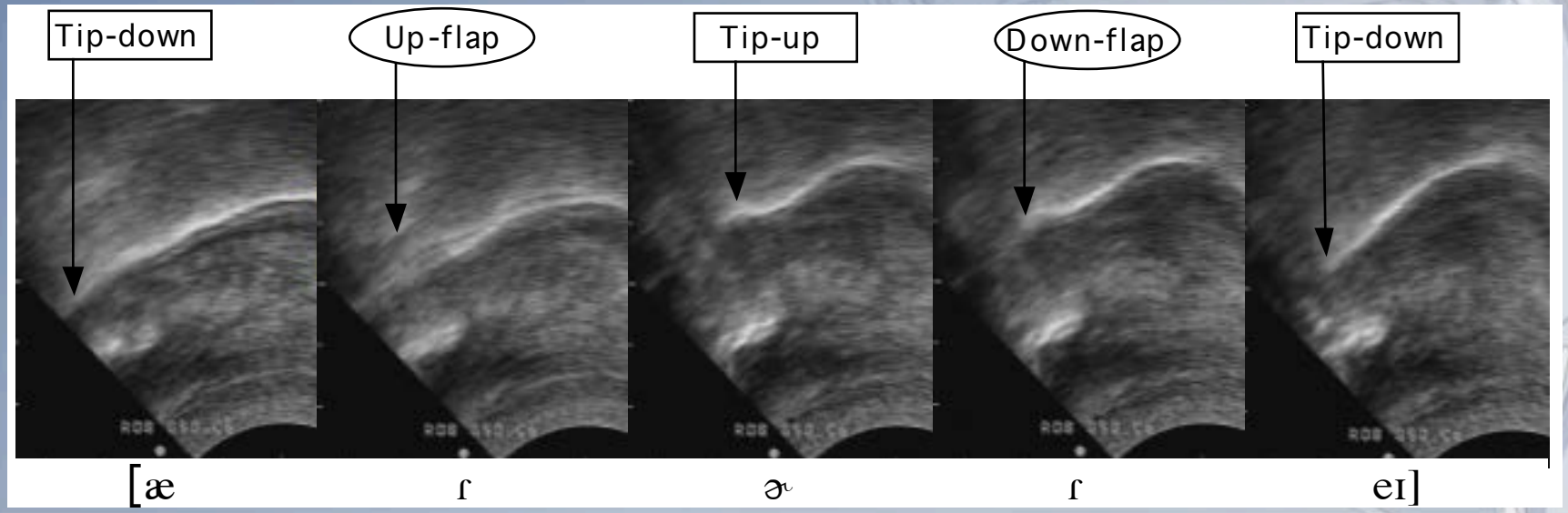
Cyclicity: Flaps

Like swallowing or walking, some modules are sequential

Flap allophones occur in context: V _ V[-stress] (e.g., *butter*, *amity*)

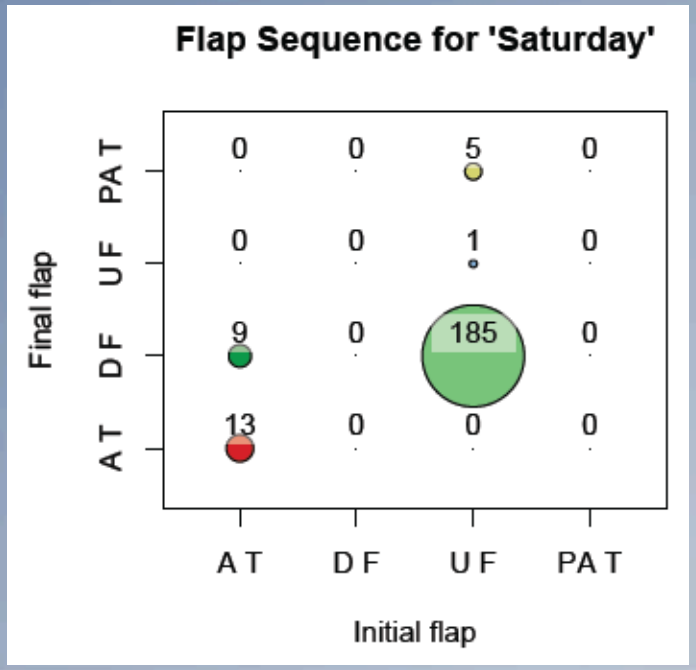
4 kinematic alternatives: **up-flap / down-flap / high tap / low tap**

Sequences get complicated...as in the word *Saturday*:

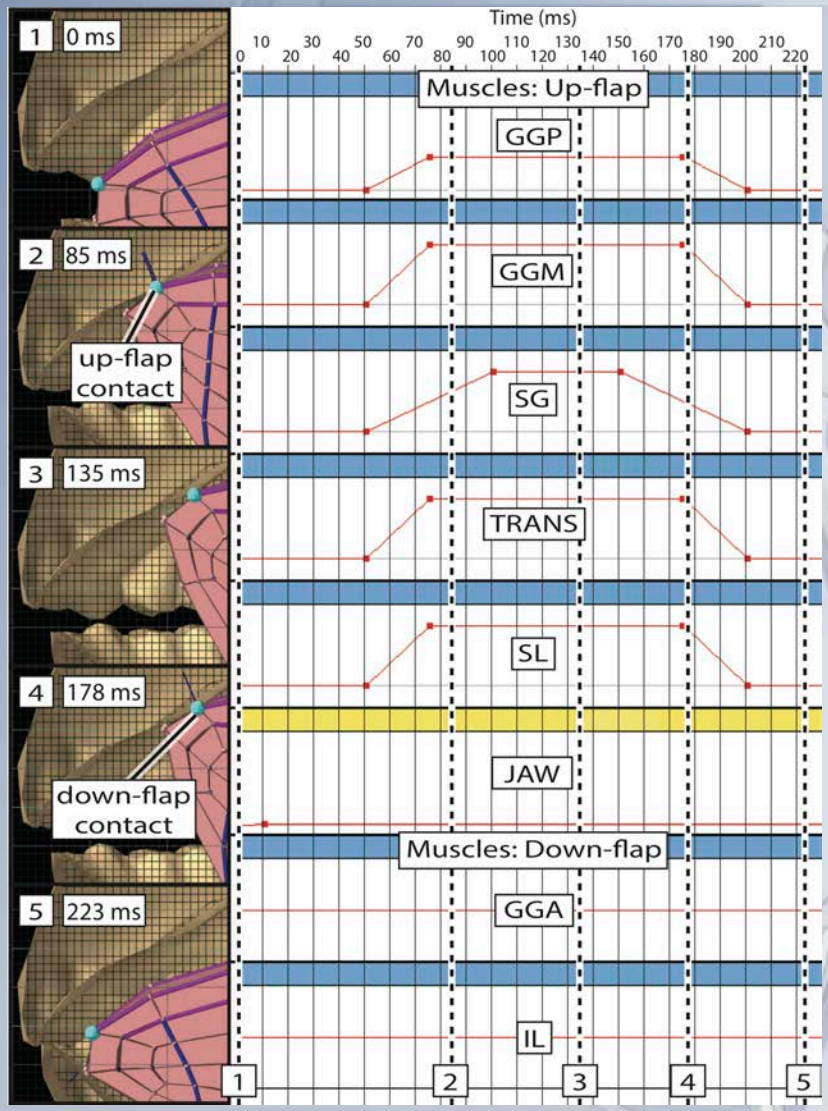


Cyclicity: Flaps

sequences of coronal stops overwhelmingly realized as ***up-down flap*** pairs



from Derrick, Gick & Stavness (2010)

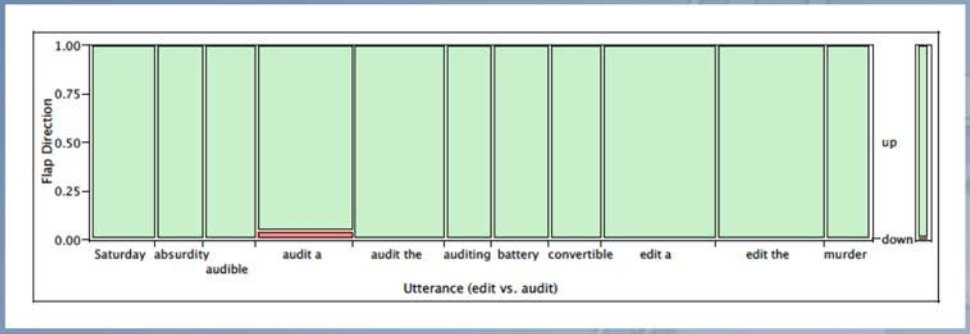


from Derrick, Stavness & Gick (2015)

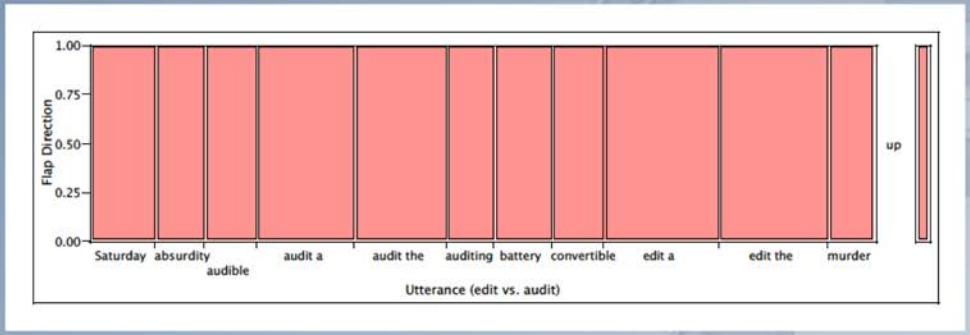
Motor Abundance (an aside)

Individual Variation in Flap Direction: Type I

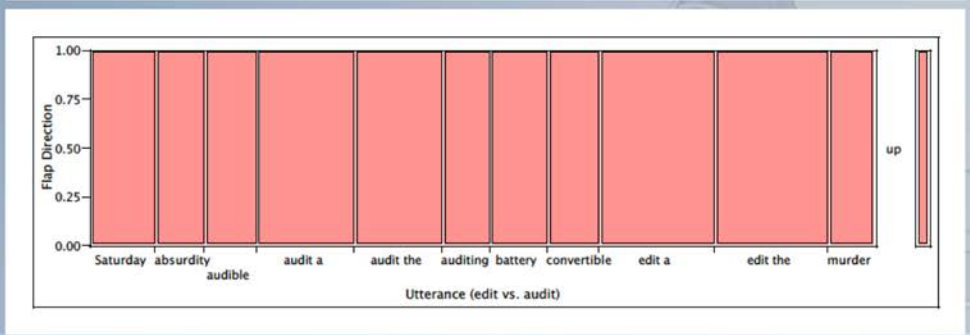
Subject AG



Subject CB



Subject MJ

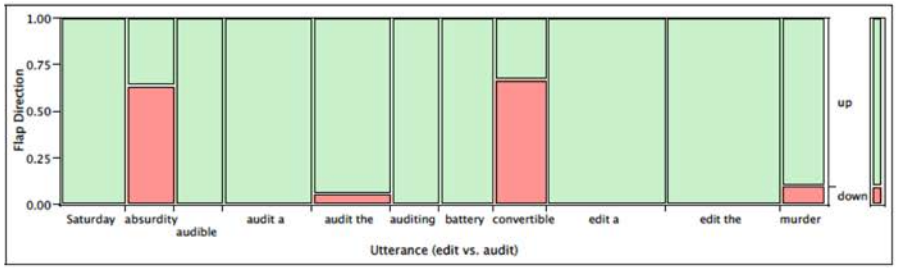


All up-flaps, regardless of conflicts

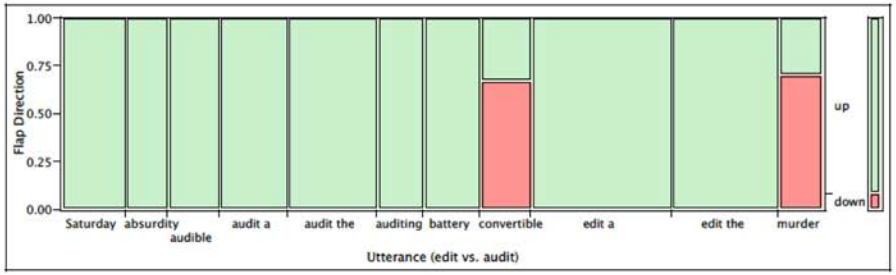
Motor Abundance (an aside)

Individual Variation in Flap Direction: Type II

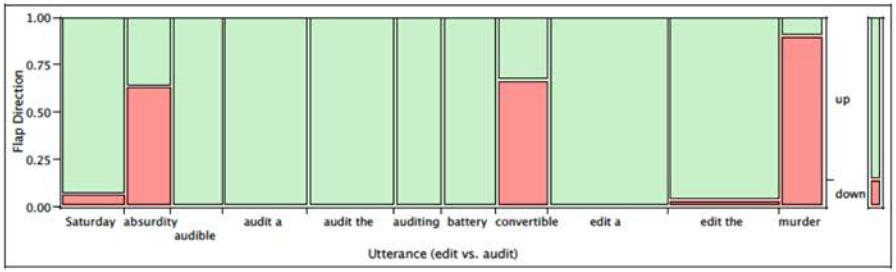
Subject JH



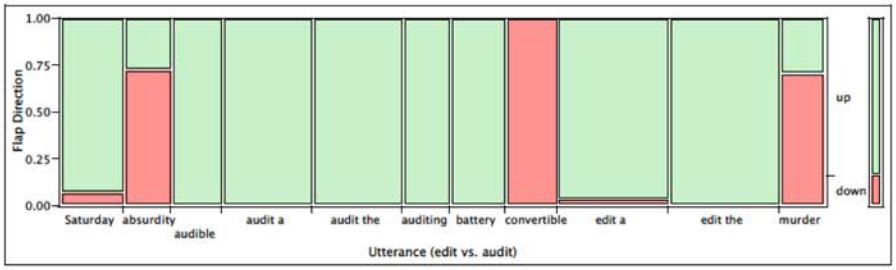
Subject KK



Subject RU



Subject SS

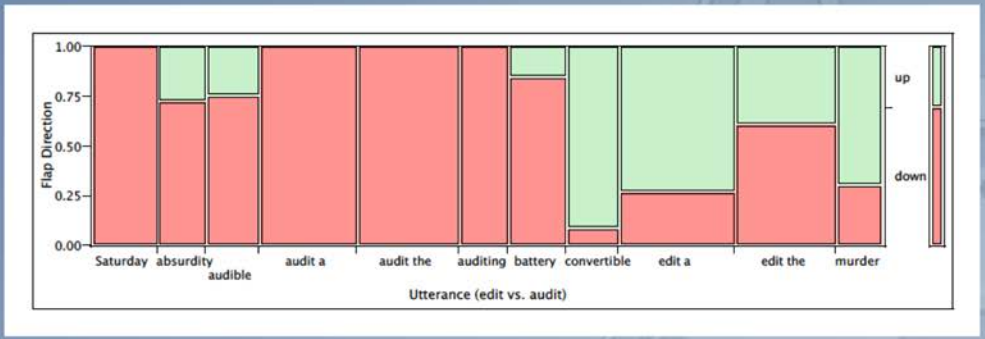


Variability of strategy in *absurdity*, *convertible* and *murder* (/r/-initial sequences)

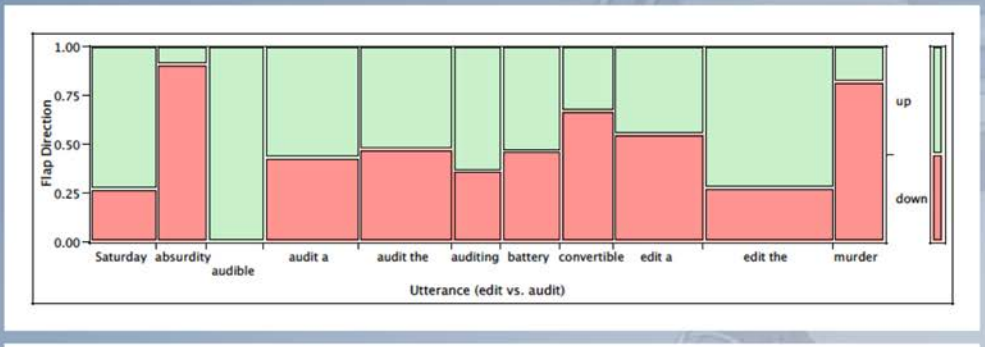
Motor Abundance (an aside)

Individual Variation in Flap Direction: Type III

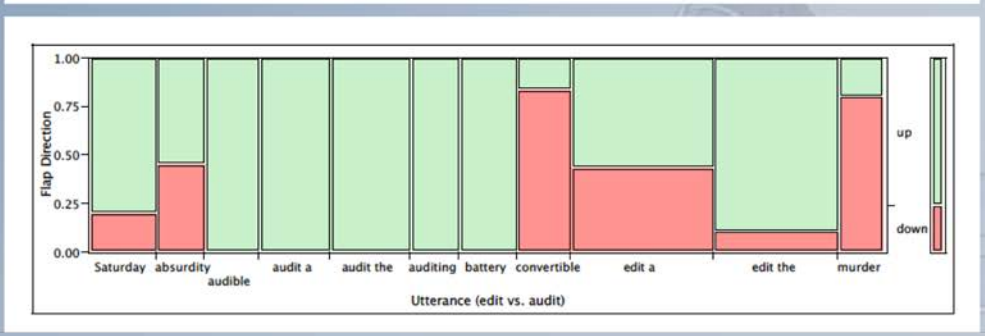
Subject CC



Subject CT



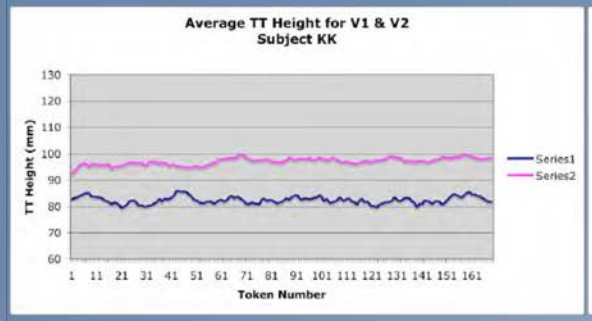
Subject MM

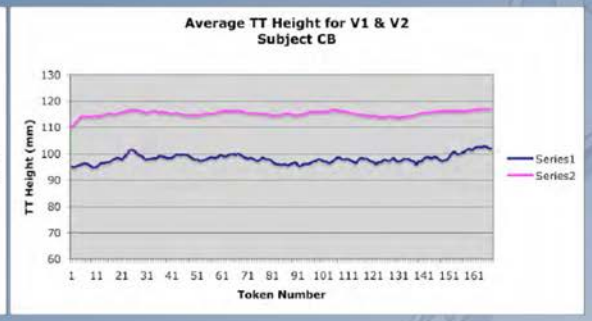


Variability of strategy in a variety of sequences

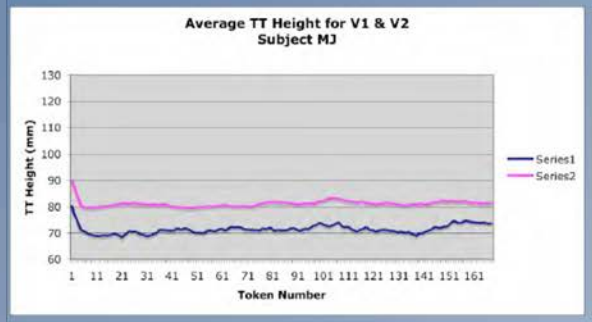
Motor Abundance (an aside)

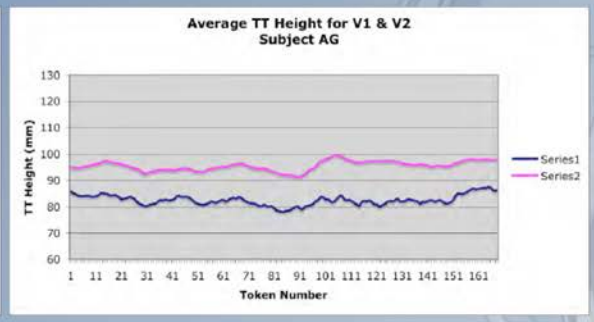
Global Patterns Across All Utterances (V1&V2)

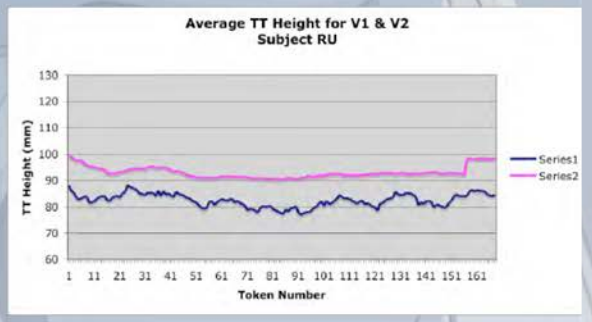


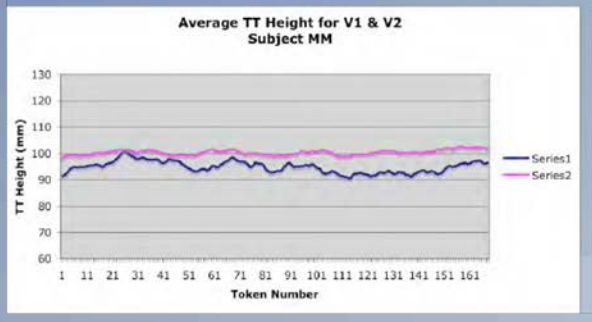


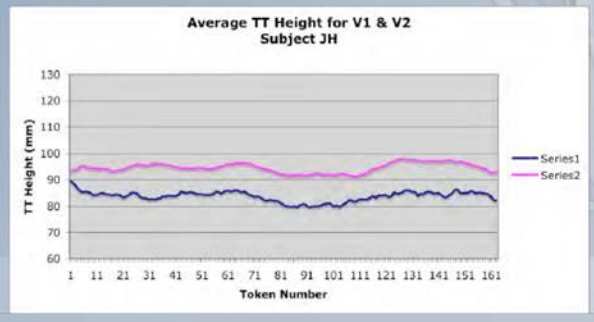
Most subjects prefer up-flaps across utterances

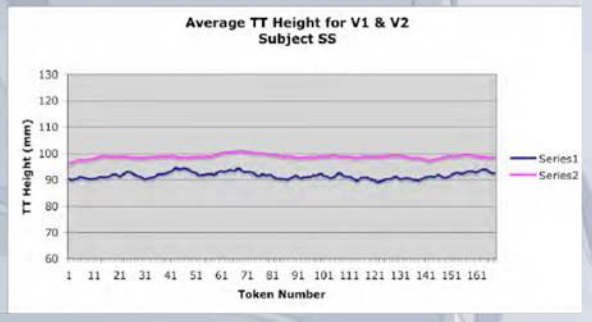








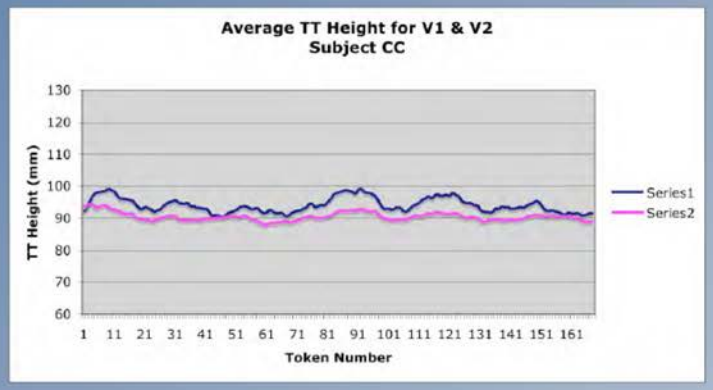


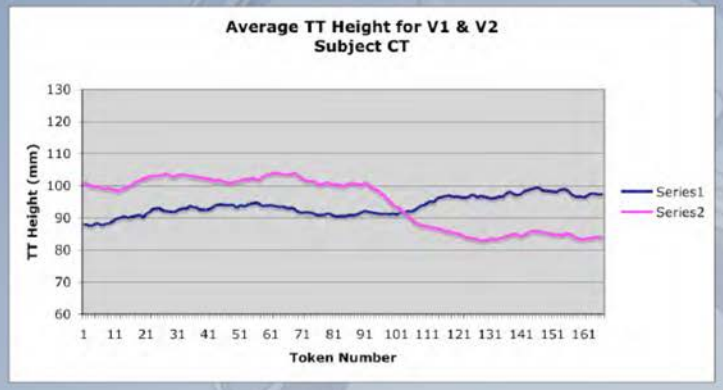


Motor Abundance (an aside)

Global Patterns Across All Utterances (V1&V2)

Global Strategy Shifts!*





(reverse of normal pattern)

*I.e., shifts apply to all following sequences
(sub-segmental level)

Sequential Devices

- *S*imilar to transient devices
  - i.e., fixed duration, but…
- *Sequence of activations over time*

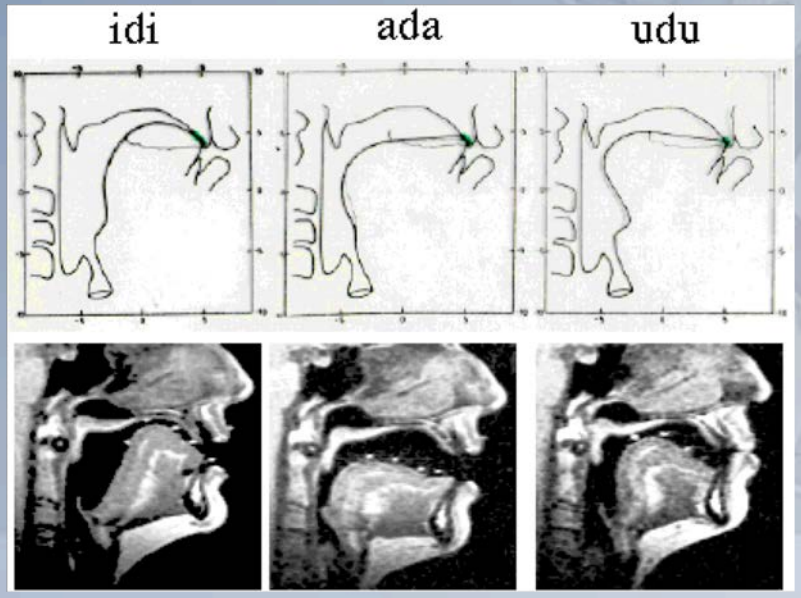
Issue: “chunk” sizes??

Coarticulation...

Generally, when a conceptually isolated speech sound is influenced by another nearby sound

OR, in more gestural terms:

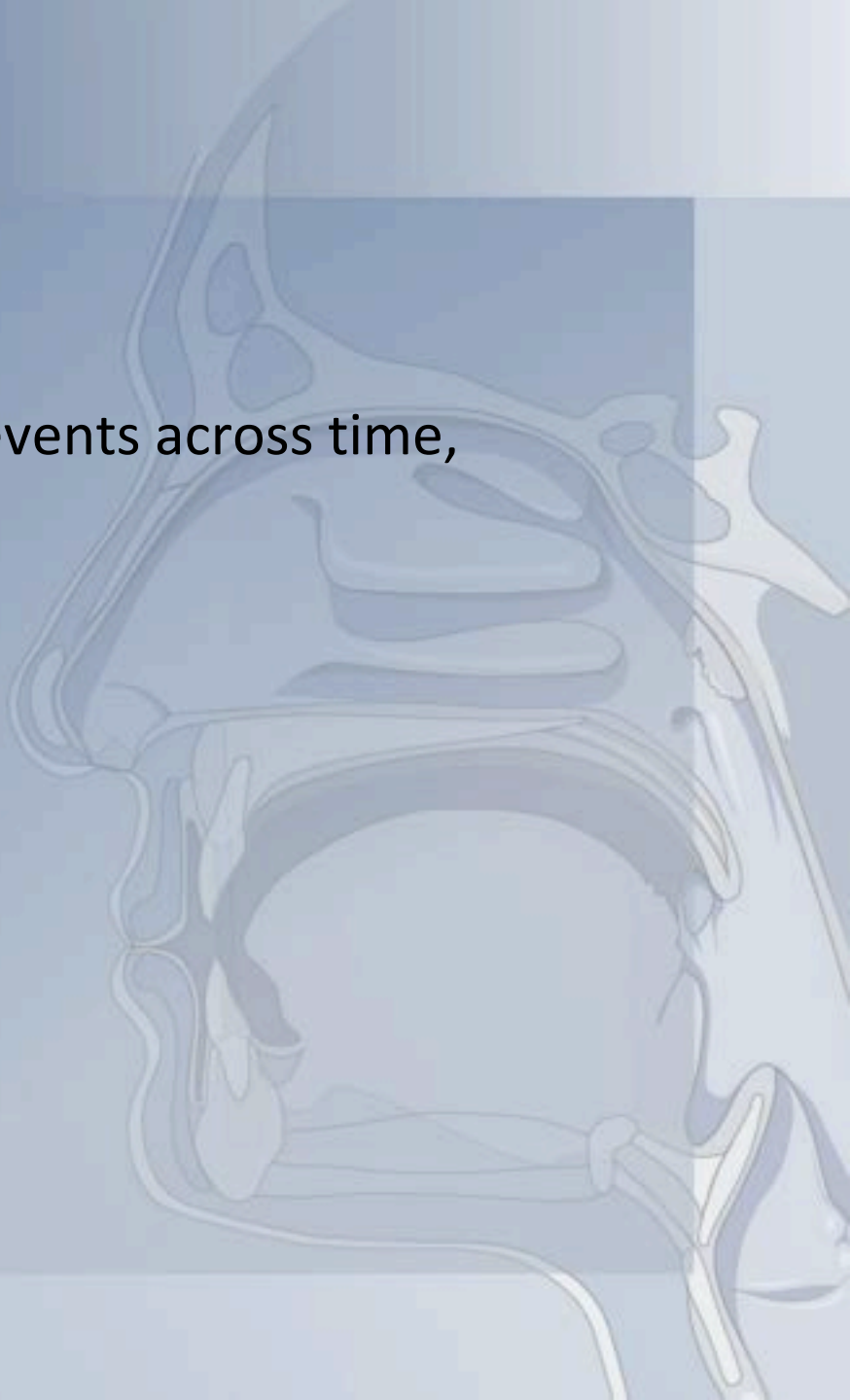
…when successive phonological units overlap in time:



Coarticulation

Once we're looking at speech events across time,

the inevitable will happen…



Coarticulation...

…when successive phonological units overlap in time

⇒ Sometimes represented as actions of multiple “parts”

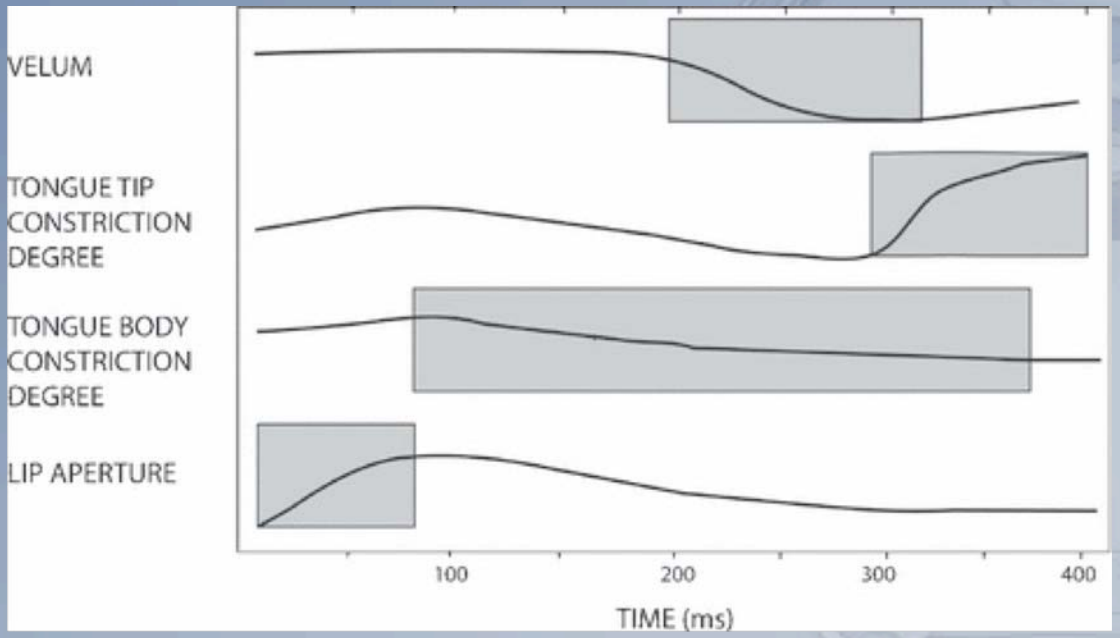


Image from Browman & Goldstein

What is coarticulation?

What gets concatenated when we speak?

What is the scope of coarticulation/concatenation?

Coarticulation is an unsolved combinatorial problem

...stacked on top of an unsolved combinatorial problem!

How many things are going on here?

Samuel L. Jackson - Pulp Fiction (1994)

How many things are going on here?

Embodied coarticulation...

People use their bodies to:

- Survive – digest, respire…
- Express/transmit cultural information
- Express/transmit emotional state
- Express/transmit language/dialect identity
- Produce/transmit specific spoken messages
- Direct attention/gaze
- Maintain head angle/posture
- Etc...

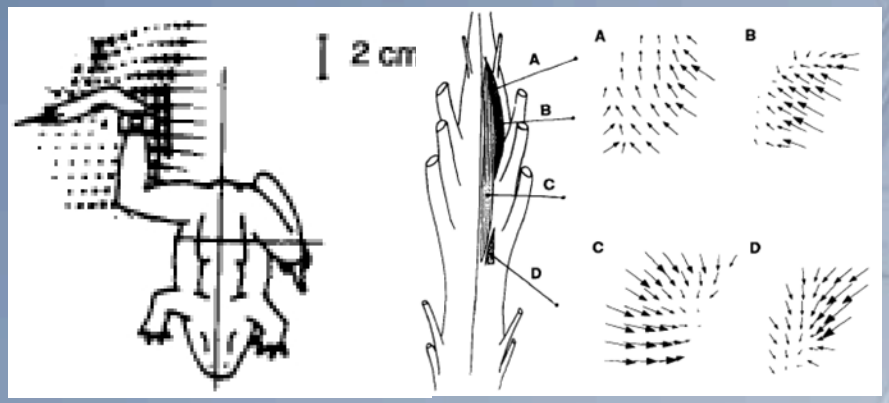
...all simultaneously!

How many things are going on here?

Samuel L. Jackson - Pulp Fiction (1994)

Neurophysiological Modules

We've long known of functional modules in motor systems



"spinalized" frogs
Bizzi et al. (*Science* 1991)

- Neural pathways corresponding to *useful* and *reliable* actions
- Solution to Berstein's dimensionality ("DOF") problem

…An important discovery: "superposition"

- overlapping modules are *additive*!

Superposition: Coarticulation

…hearkens back to Joos' (1948) foundational "Overlapping Innervation Wave Theory"

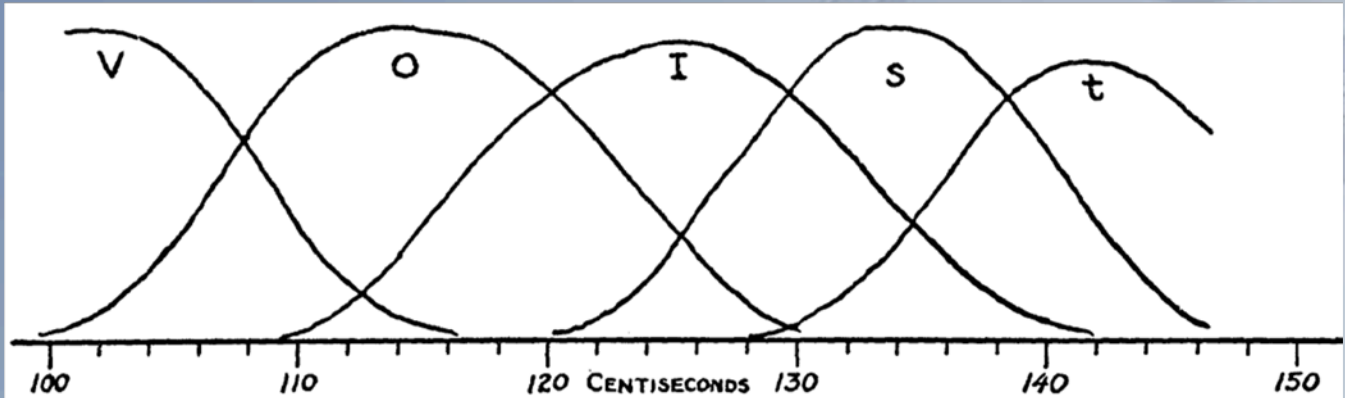
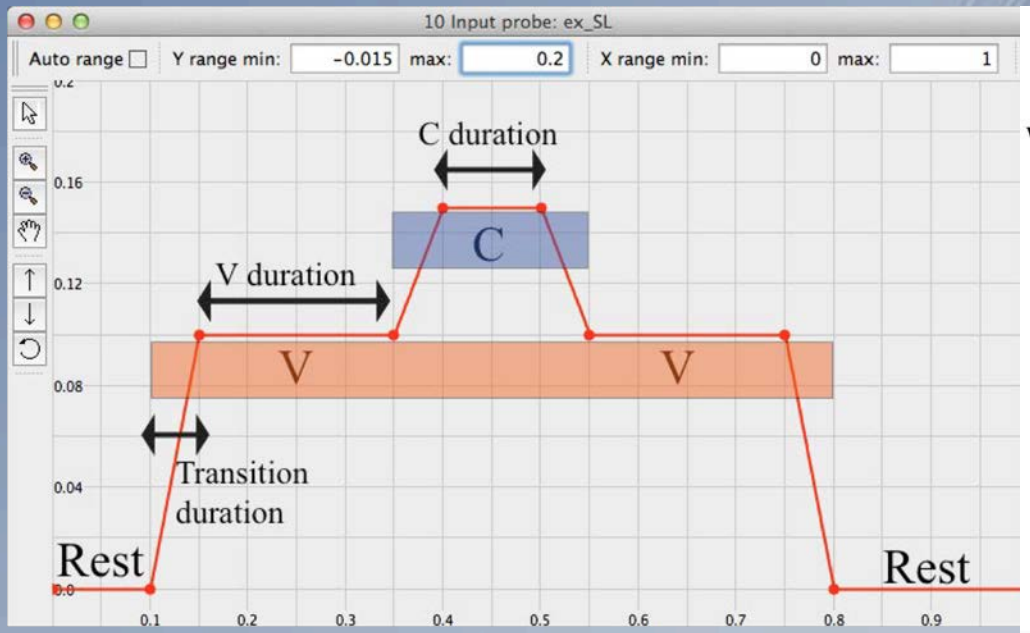


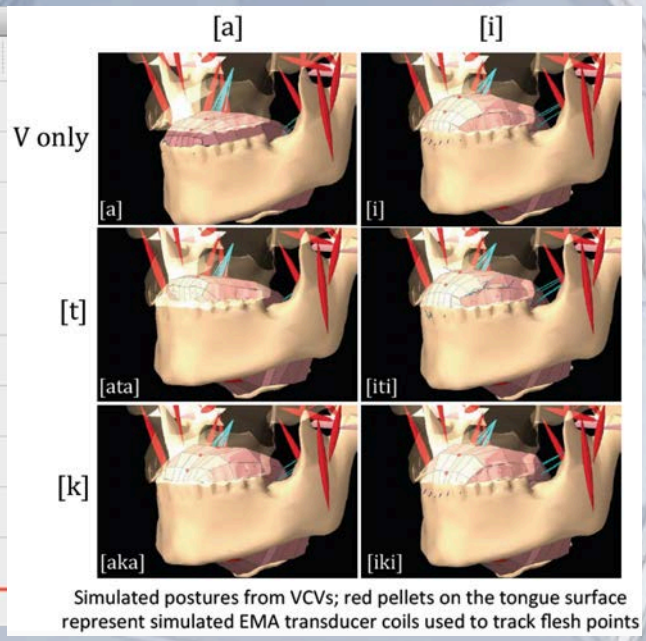
FIG. 38. OVERLAPPING INNERVATION WAVES

Can we simulate speech production results of coarticulation…with *no model of coarticulation*?

Superposition: Coarticulation

Simulation of Recasens and Espinosa (2009) EMA results - from Gick et al (2013)

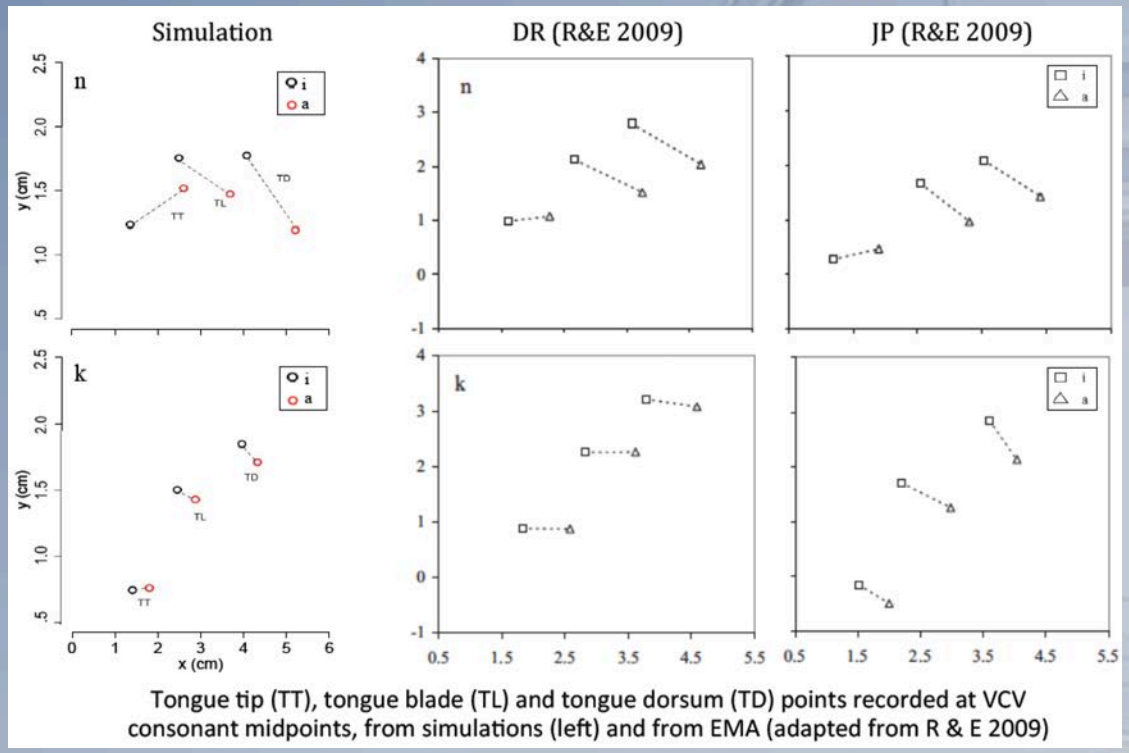




Simulated postures from VCVs; red pellets on the tongue surface represent simulated EMA transducer coils used to track flesh points

Superposition: Coarticulation

Simulation of Recasens and Espinosa (2009) EMA results



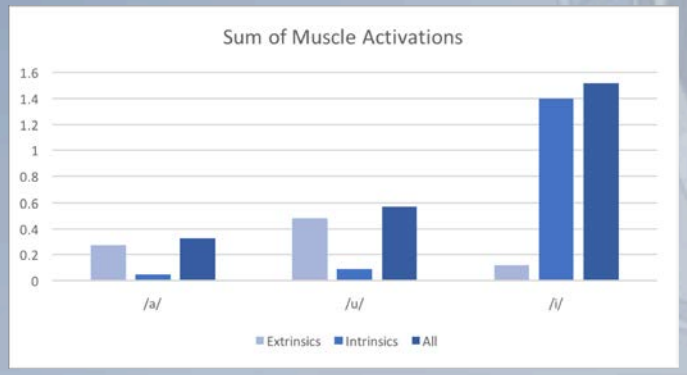
Tongue tip (TT), tongue blade (TL) and tongue dorsum (TD) points recorded at VCV consonant midpoints, from simulations (left) and from EMA (adapted from R & E 2009)

Superposition: Coarticulation

Recasens and Espinosa (2009) also find that coarticulatory "resistance" and "aggressiveness" are positively correlated…

- Palatals are both more resistant and more aggressive!
- Indeed, they are the same thing in body space...

…*stiffness* (aka activated muscles):



(Gick et al., Interspeech 2017)

Superposition: Coarticulation

Our approach treats all coarticulation as *superposition*…

- Survive – digest, respire…
- Express/transmit cultural information
- Express/transmit emotional state
- Express/transmit language/dialect identity
- Produce/transmit specific spoken messages

ALSO speech-specific cases of coarticulation…

- local (canonical “coarticulation”)
- non-local (long-distance “harmony”)
- global (articulatory setting)

…with *no model* (except a pretty good model of the body)

Superposition: Harmony as "Non-local Coarticulation"

$CV_iC + CV_jC > CV_jCCV_jC$

Gafos (1996) proposed a kinematic explanation:
=> differences in *tongue shape* maintained
across intervening sounds
(long-distance coarticulation)

but this explanation breaks down in some cases…

Superposition: Harmony as "Non-local Coarticulation"

Tahltan sibilant harmony (Shaw, 1991)

(22)	a.	/s/	*ɛdɛdɛθ̄du : θ*	'I whipped myself'
		/s/	*taθ̄tθaɬ*	'I'm dying'
		/s/	*xaʔɛθ̄tʼaθ*	'I'm cutting the hair off'
		/θ/	*dɛs̄itʼʌs*	'we are walking'
		/θ/	*nis̄itʼa : ts*	'we got up'
		/θ/	*mɛʔɛs̄itʼotš*	'we are breast-feeding'
	b.	/s/	*noʔɛdɛ : š̄ɬɛdǯi*	'I melted it over and over'
		/s/	*yaš̄tɬʼɛtš*	'I splashed it'

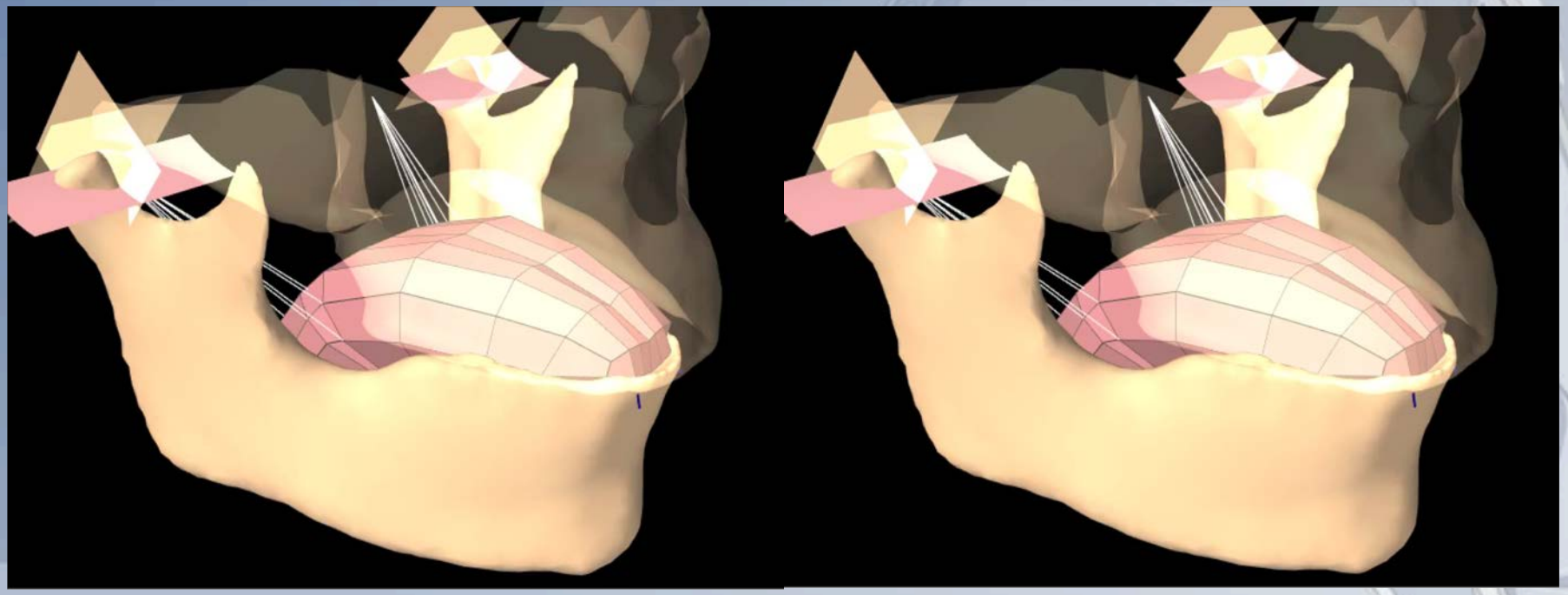
Harmony passes over *coronal stops* and *lateral fricatives*!

Superposition: Harmony as "Non-local Coarticulation"

Gick, Whalen, Shaw & Stavness (2013) use simple simultaneity/superimposition – the body works out the averaging

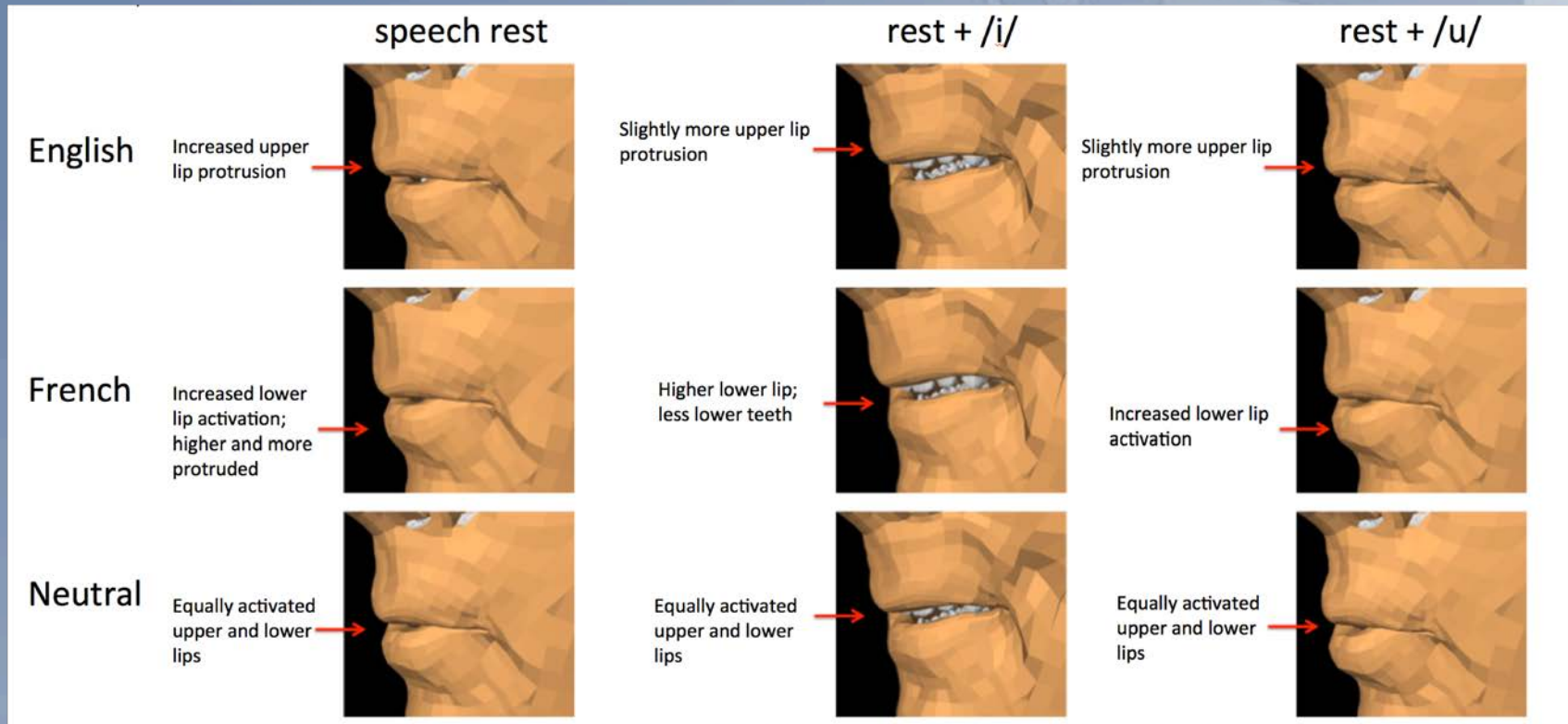
Superimposed coronal stop

Superimposed lateral



…same solution for *articulatory settings/PSPs/ISPs* => tonic activation!

Articulatory Setting as "Global Coarticulation"



=> Native English expert listeners labeled English correctly 2/3 of the time (at chance labelling French)

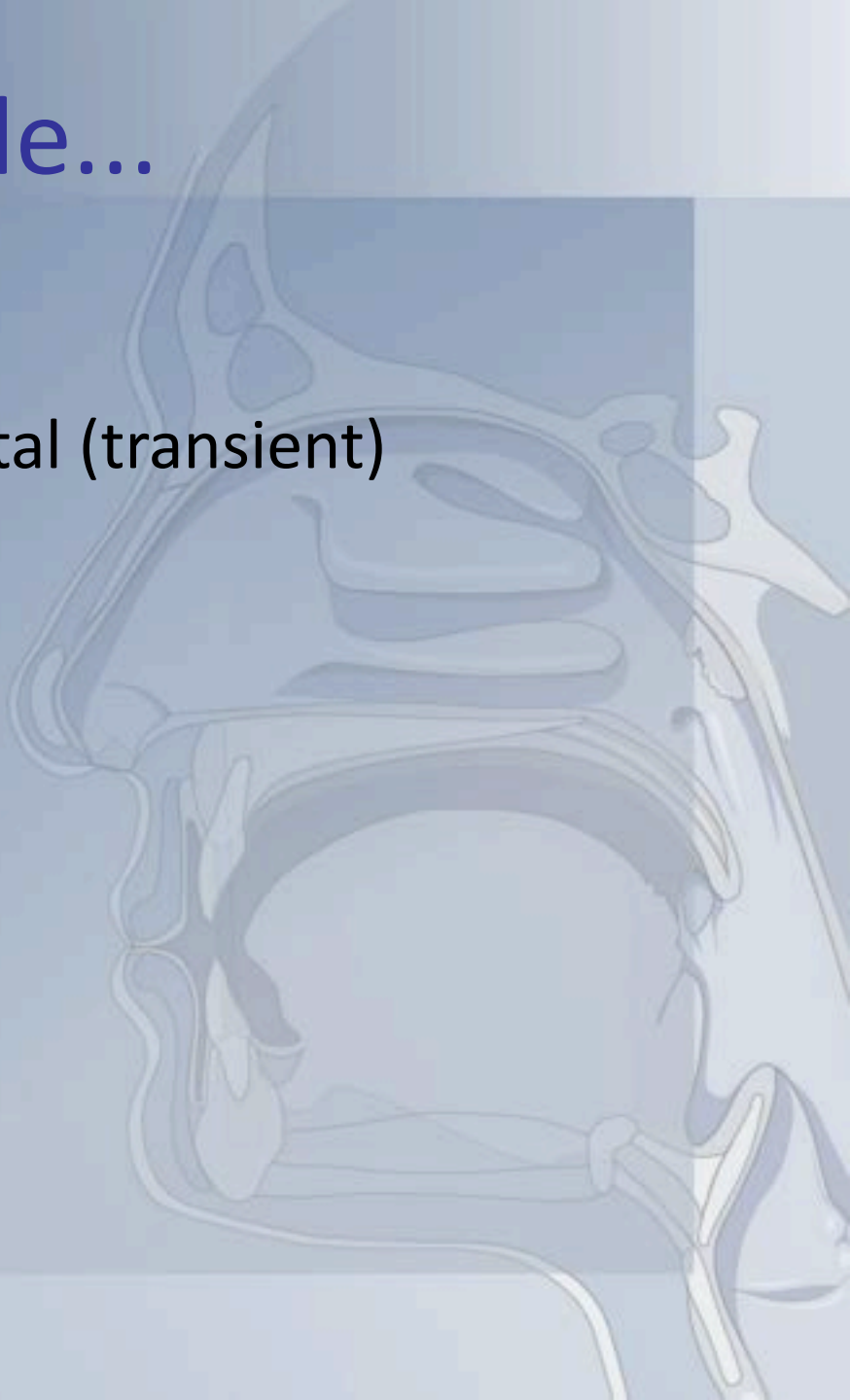
=> Listeners labeled [i] correctly 70% of the time (at chance on rest and [u])

=> Neutral tokens labeled French 80% of the time

From Gick, Chiu, Roewer-Despres, Schellenberg & Stavness (2016)

Another example...

- Smile (tonic) + labiodental (transient)



Thank you!