

# **Task Dynamics Application (TADA) for articulatory speech synthesis**

**Jaekoo Kang**

Speech-Language-Hearing Sciences  
CUNY Graduate Center  
Haskins Laboratories

2020-10-19

# Plans

## Tutorial 1 (Oct. 19)

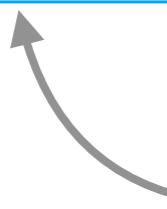
### Matlab and TADA basics

- Part 1. Matlab basics
- Part 2. Intro to TADA

## Tutorial 2 (Oct. 26)

### Synthesis using TADA

- Vowel gestures & timing.
- Consonant gestures & timing.



Today!

# Acknowledgements

- This tutorial slides were created based on Articulatory Phonology course taught by Hosung Nam (2014).
- Additional materials were reused and modified from the TADA manual and slides from Hosung Nam and Louis Goldstein.

# **Part 1. Matlab basics**

# Overview

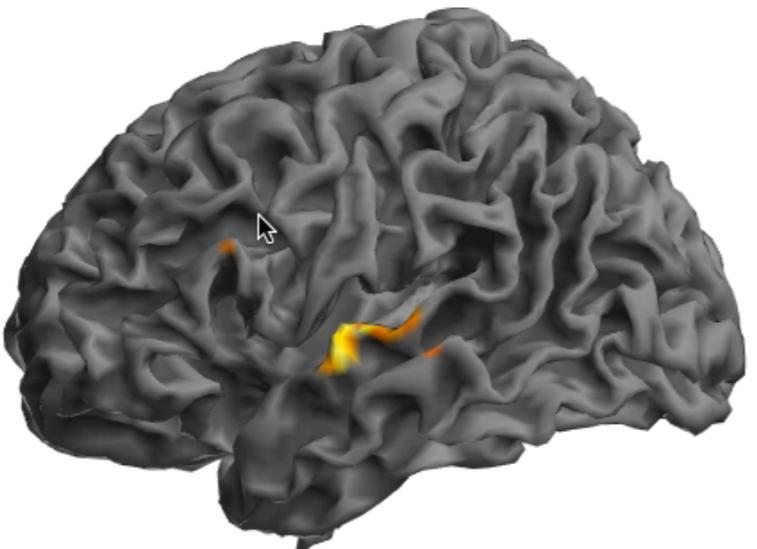
- Matlab interface
- Commands
- Variables & operations
- Loops
- Plotting
- help

# Matlab

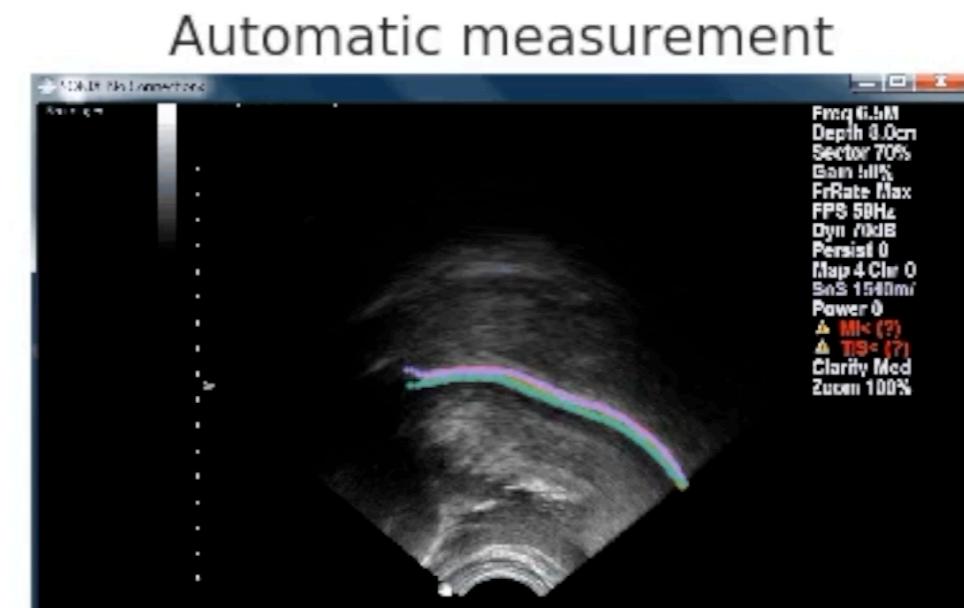
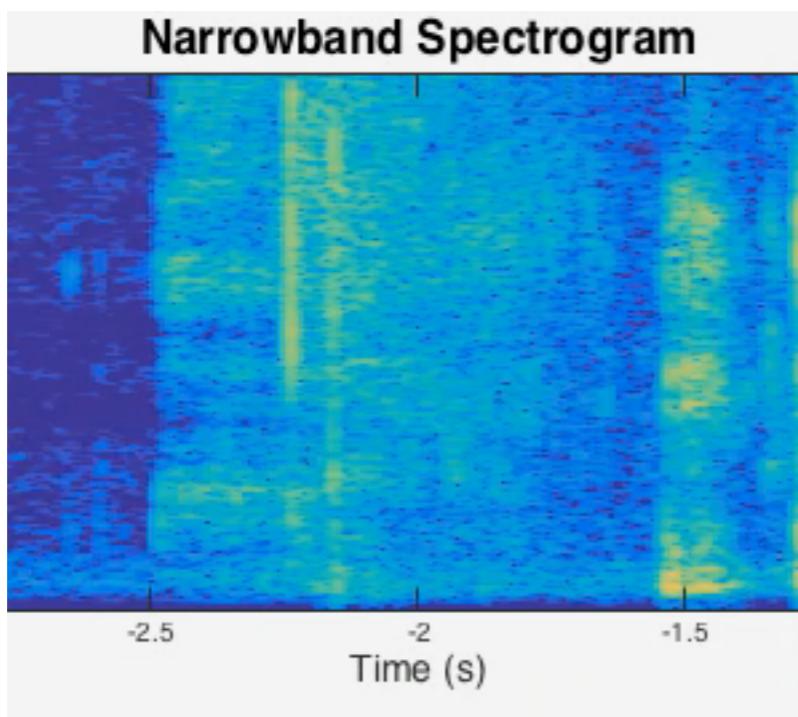
- A programming language designed for scientists and engineers.
- Pros:
  - Easy to use.
  - Useful built-in functions.
  - Documentations.
- Cons:
  - Pricing.
  - Lack of flexibility.



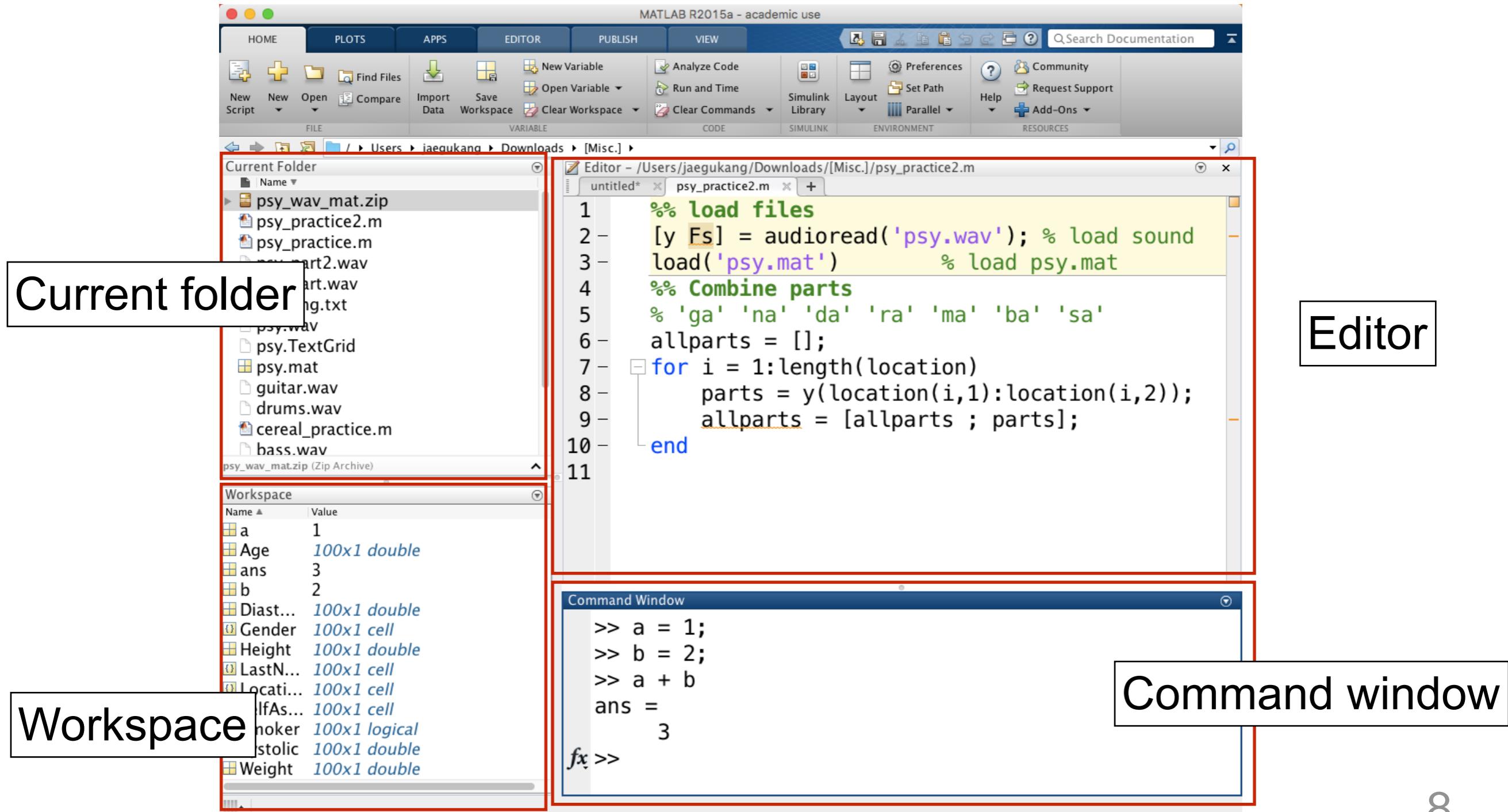
# Matlab for SLHS?



● Alex and Tim are lost. Alex's wife recently passed away and he couldn't find the motivation to live a normal life. Tim has been charged with embezzlement and is going to prison soon. The thought of that terrified him. They decided to take a ride down the river rapids at Cataract Canyon. A river only goes one way. If you are lost, that is good. Alex has never rafted on this river before, but Tim has many times. Tim tells Alex that



# User interface



# Giving commands

- Command window:

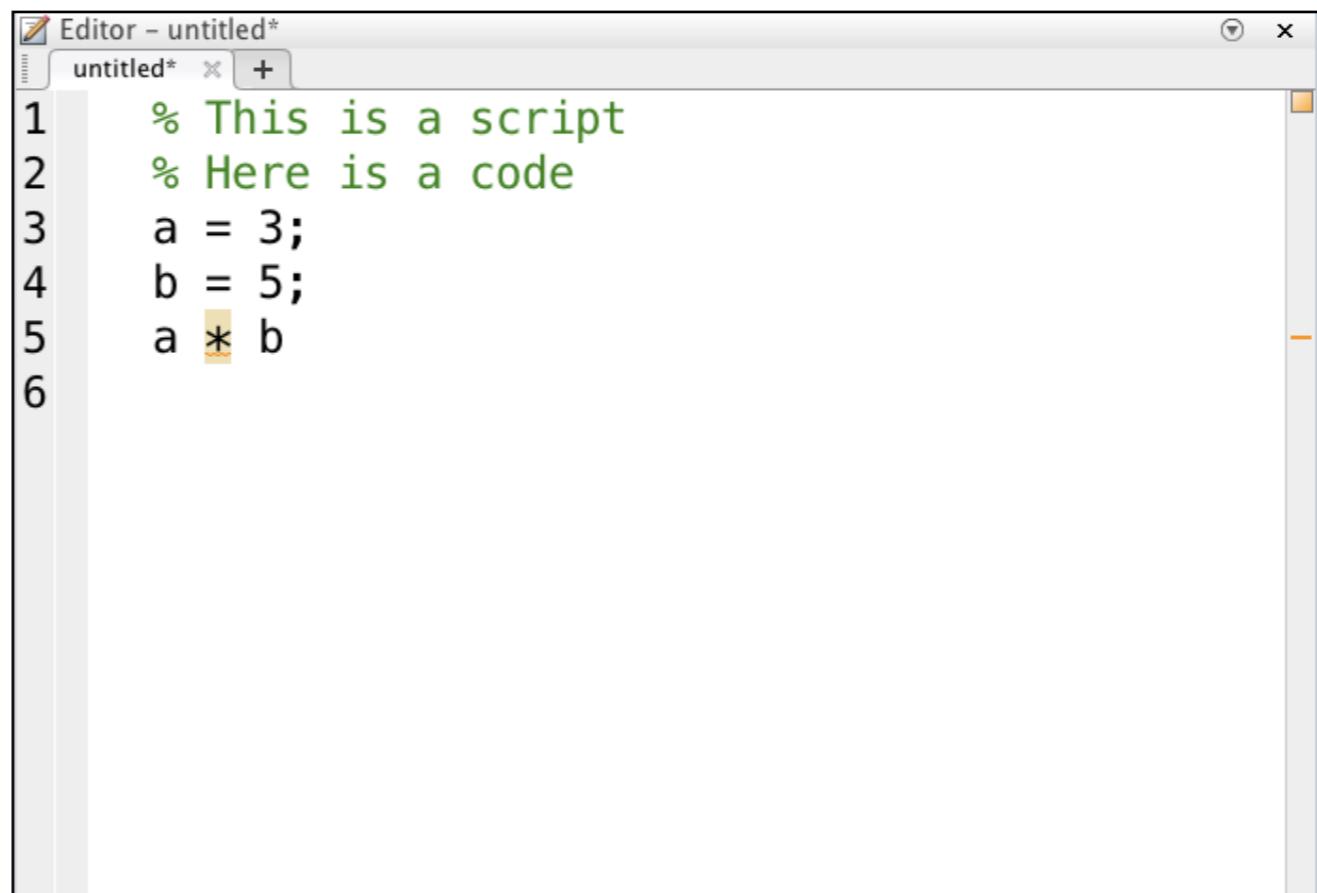
- simple commands.



The screenshot shows the MATLAB Command Window with the following text:  
>> a = 3;  
>> b = 5;  
>> a \* b  
ans =  
15  
*fx* >> |

- Editor window:

- complex commands.
- procedure.



The screenshot shows the MATLAB Editor window titled "Editor - untitled\*" with the following code:  
1 % This is a script  
2 % Here is a code  
3 a = 3;  
4 b = 5;  
5 a \* b  
6

# Variable types

The image shows the MATLAB interface with the Command Window and Workspace panes. The Command Window displays the following code:

```
>> a = 3;
>> b = 'hello!';
>> c = [1, 0.8];
>> d = [1 2; 3 4];
>> e = {0.2, 'k'};
```

The Workspace pane shows the variables and their values:

Name	Value
a	3
b	'hello!'
c	[1,0.8000]
d	[1,2;3,4]
e	1x2 cell

Number  
Character  
Vector  
Matrix  
Cell

# Basic operations

1      Command Window

```
>> 4/3
ans =
1.3333
```

2

```
>> 9^2
ans =
81
```

3

```
>> A = [1 2;3 4;5 6];
>> B = [9 6;-1 4;1.2 3];
>> A + B
ans =
10.0000    8.0000
2.0000    8.0000
6.2000    9.0000
```

4

```
>> A - B
ans =
-8.0000   -4.0000
4.0000      0
3.8000    3.0000
```

5

```
fx >> A * B??
```

- Operators:

+ - \* / ^ . '

# Loops

1

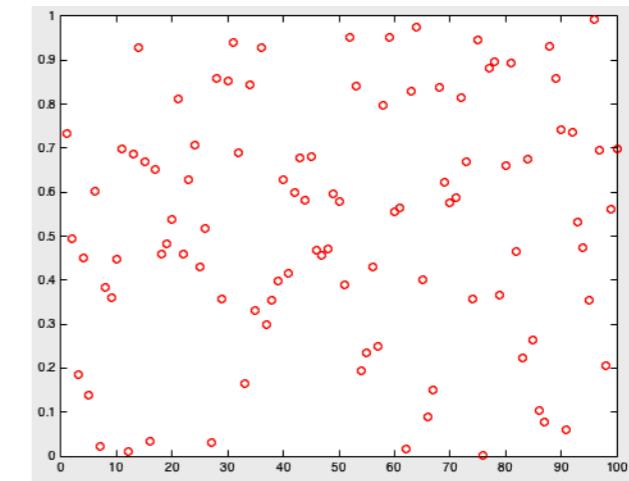
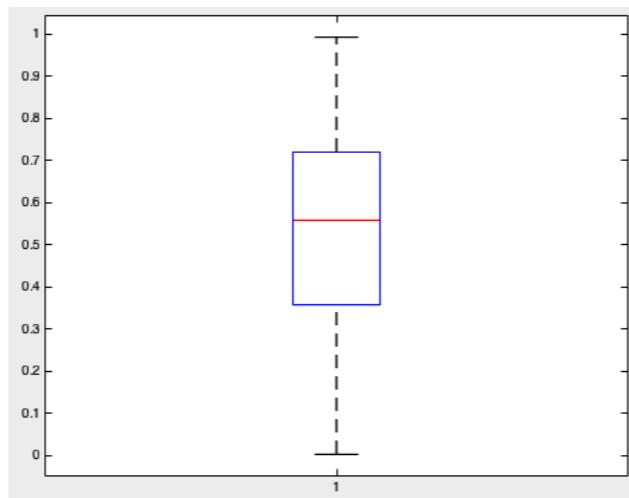
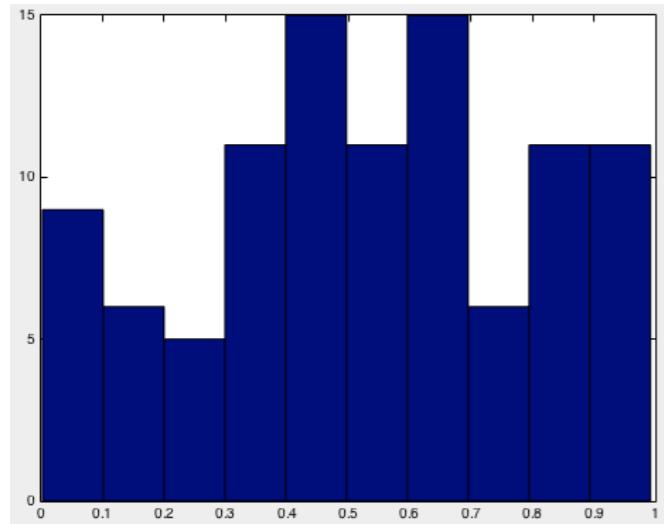
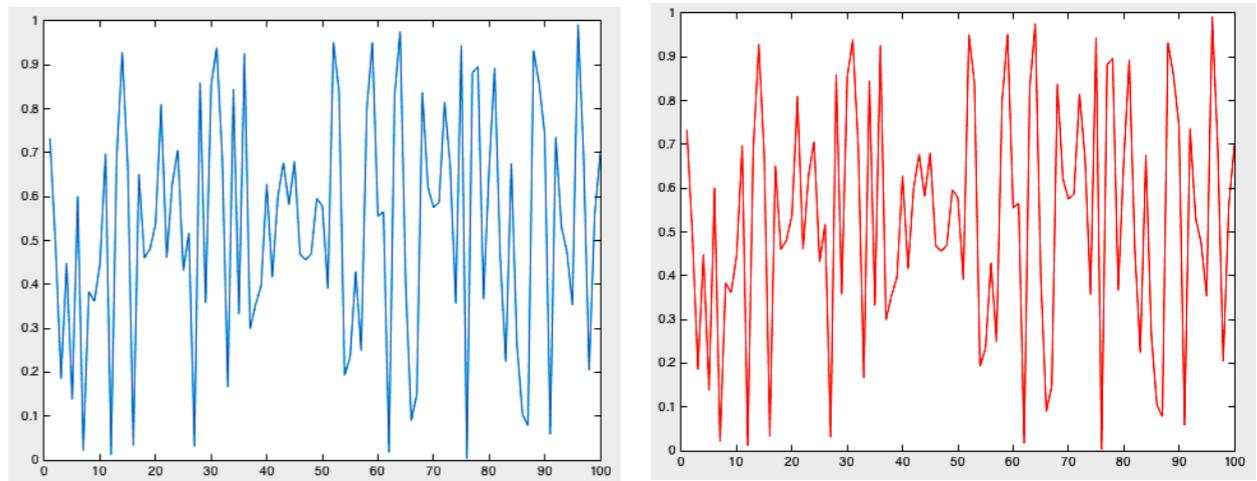
```
1   for i = 1:10
2       if i < 5
3           a(i) = i^2;
4       end
5   end
6
```

2

```
1 pitch=[105,200,210,108,101,99,180,220];
2 % Q) mean pitch?
3 % Q) SD of pitch?
4 % Q) pitch higher than 130?
5 %     -- using for-loop
6 %     -- without for-loop
7
```

# Plotting

```
1 x = rand(1,100);  
2 plot(x)  
3 plot(x, 'r')  
4 plot(x, 'ro')  
5 hist(x)  
6 boxplot(x)
```



# Handling a table data

	1 Speaker	2 Vowel	3 Duration	4 F1	5 F2
1	F01	IY1	0.1497	416.5000	2.5469e+03
2	F01	UW1	0.0499	387.4000	1.4094e+03
3	F01	AE1	0.0998	751.7000	1.7022e+03
4	F01	AA1	0.1297	776.2000	1.2477e+03

• • •

- Single speaker: F01
- 4 vowels: IY1, UW1, AE1, AA1 (/i, u, æ, a/)
- 2 categorical variables: Speaker, Vowel
- 3 continuous variables: Duration, F1, F2

# Handling a table data

	1 Speaker	2 Vowel	3 Duration	4 F1	5 F2
1	F01	IY1	0.1497	416.5000	2.5469e+03
2	F01	UW1	0.0499	387.4000	1.4094e+03
3	F01	AE1	0.0998	751.7000	1.7022e+03
4	F01	AA1	0.1297	776.2000	1.2477e+03

Q) Distribution of vowel IY1 and AE1?

Q) Formant space of speaker F01?

# Handling a table data

Q) Distribution of vowel IY1 and AE1?

```
1 % Read a csv file
2 T = readtable('vowel.csv');
3 % Make 'Vowel' column as categorical variable
4 T.Vowel = categorical(T.Vowel);
5 % Make histogram
6 histogram(T.Duration(T.Vowel=='IY1'));
7 hold on
8 histogram(T.Duration(T.Vowel=='AE1'));
9 legend('IY1', 'AE1')
```

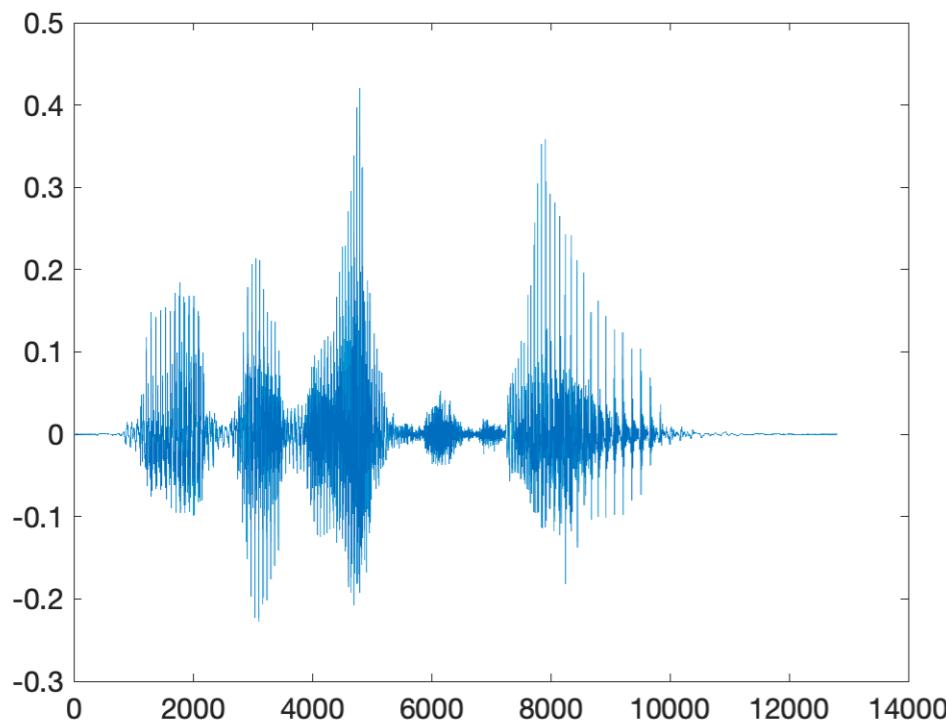
# Handling a table data

Q) Formant space of speaker F01?

```
1 % Read a csv file
2 T = readtable('vowel.csv');
3 % Make 'Vowel' column as categorical variable
4 T.Vowel = categorical(T.Vowel);
5 % Plot F1–F2
6 plot(T.F2, T.F1)
7 plot(T.F2, T.F1, 'o')
8 % Use 'set' command or property inspector!
```

# Handling sound files

```
1 % Loading a sound file
2 [y, sr] = audioread('sound.wav');
3
4 % Playing a sound file
5 soundsc(y, sr)
```



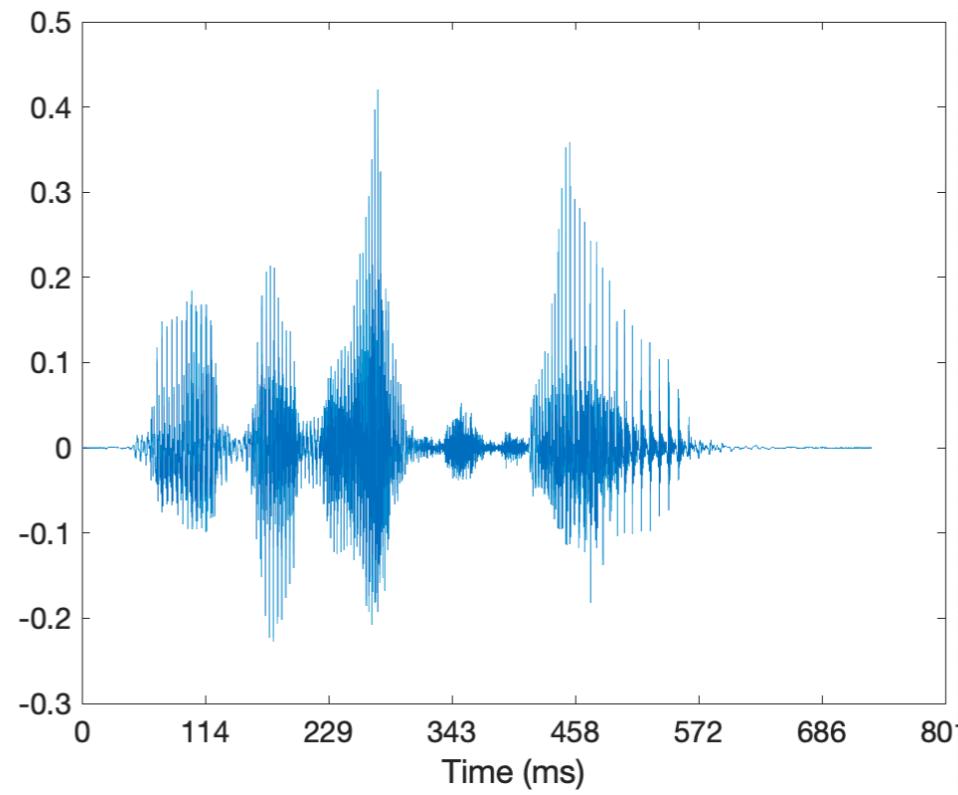
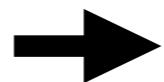
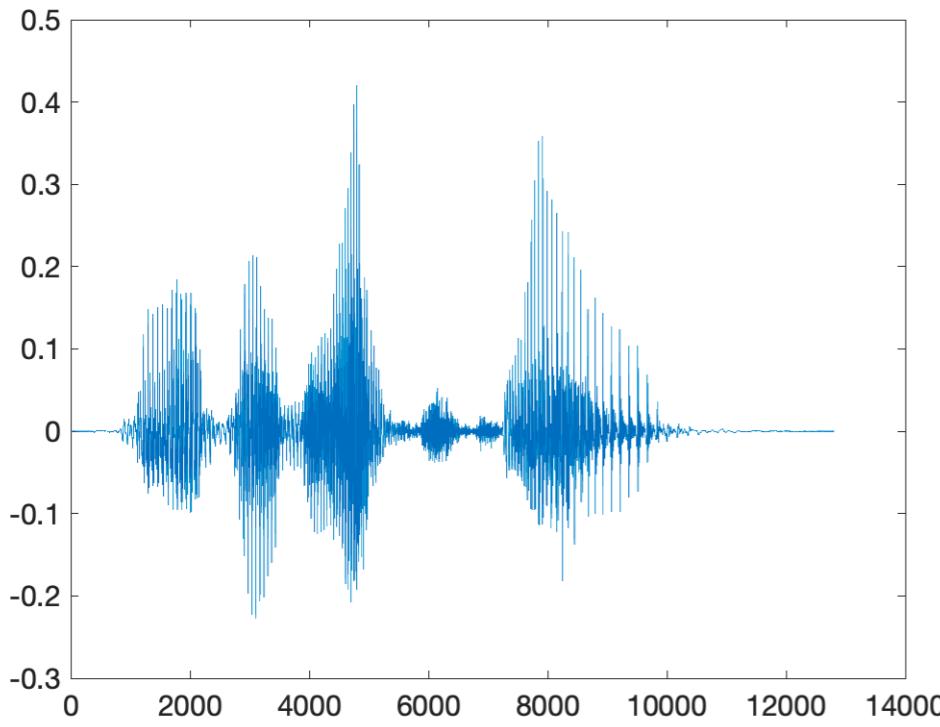
# Handling sound files

- ```
1 %% Q) How many channels are there?  
2 % ==> Check the size of the signal  
3  
4 %% Q) What is the length of the sound?  
5 % ==> Use 'length' function  
6  
7 %% Q) What is the duration of the sound?  
8 % ==> Use the sampling rate  
9  
10 %% Q) Decrease the volume by 1/2?  
11
```

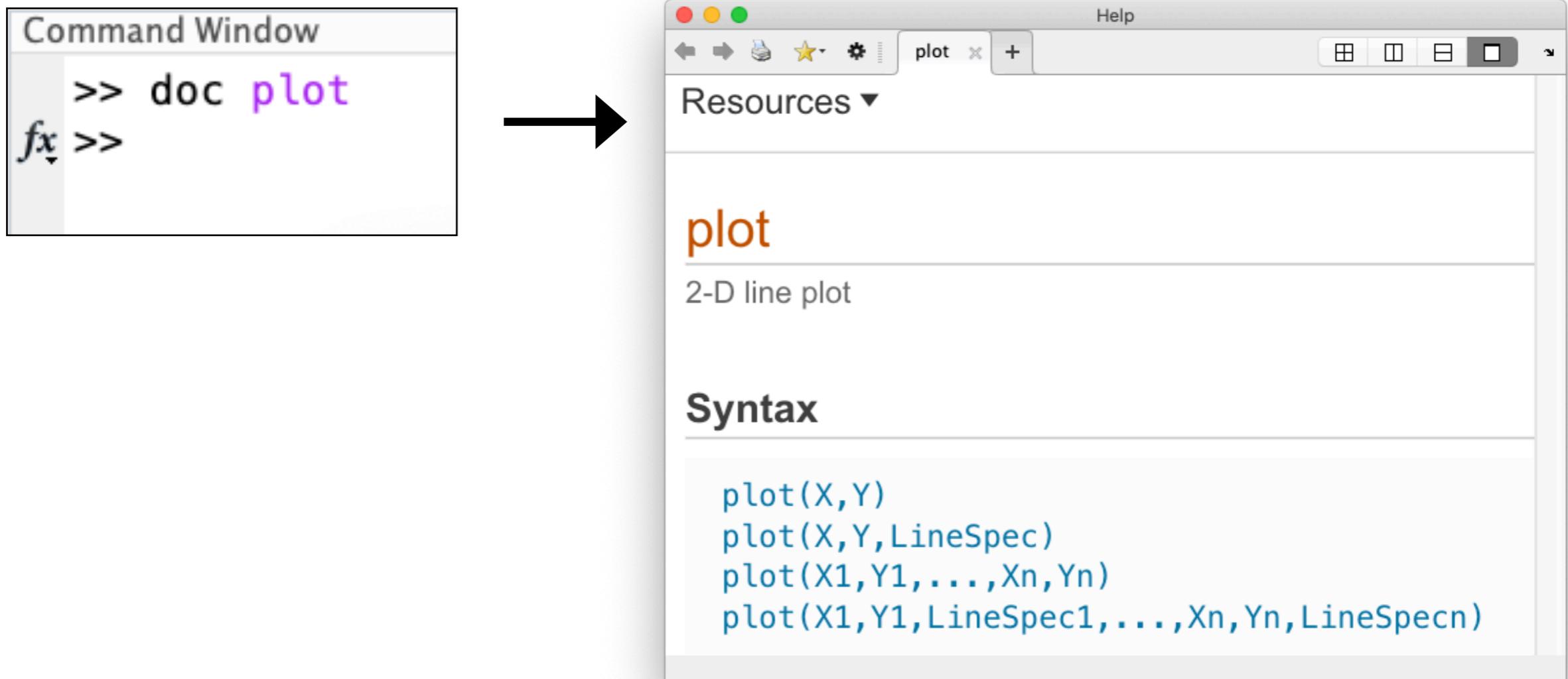
# Handling sound files

```
1 %% Plot the waveform  
2 plot(y)  
3 plot(y, 'o')
```

```
4  
5 %% Change the x tick labels  
6
```



# Help



- <https://www.mathworks.com/matlabcentral/>
- <https://stackoverflow.com/>

# Other topics

- Data input-output (e.g., images, sounds, tables etc.)
- Optimization
- Interactive visualization
- Graphic User Interface (GUI) programming
- Toolboxes (e.g., psych toolbox, EEGLAB, SPM, Eyelink etc.)

# Summary

- Matlab is useful for scientific computing and data analysis.
- Different variable types/operations are supported with various (and useful) built-in functions.
- Functions and tools are well documented.

# Questions?

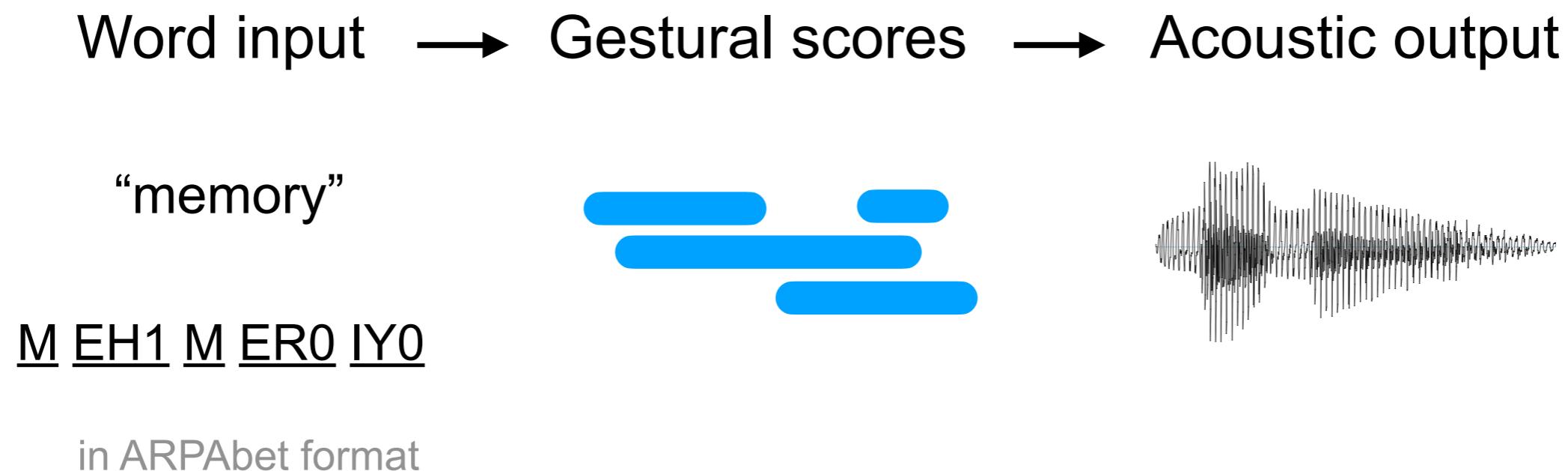
# **Part 2. Intro to TADA**

# Overview

- TADA
- Dynamical system
- Articulators and tasks
- Structure of TADA
- Practice

# TADA

- TADA (TAsk-Dynamics Application) is a dynamical model for simulating gestural structure of speech and generating acoustic output. Goldstein et al., 2006; Nam et al., 2004

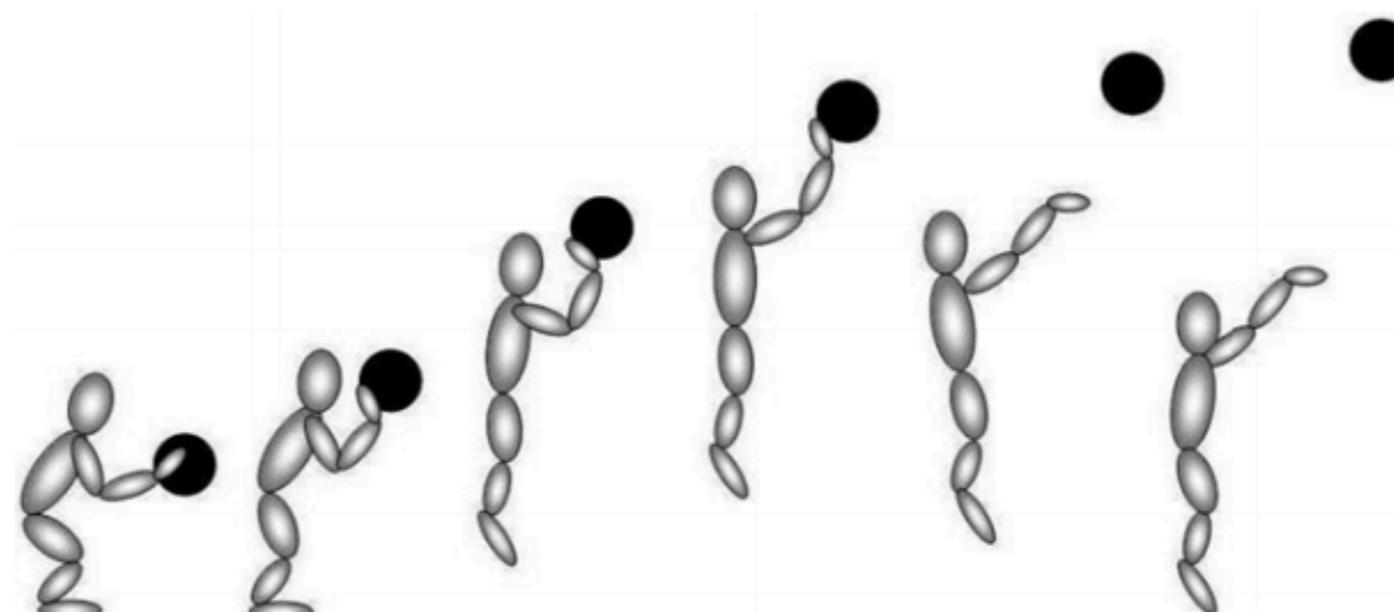


# TADA

- TADA is written in Matlab. The current version is v0.983, maintained by Hosung Nam and Jaekoo Kang.
- Developers (alphabetical)
  - Catherine Browman
  - Louis Goldstein
  - Hosung Nam
  - Philip Rubin
  - Elliot Saltzman
  - Mark Tiede

# Dynamical system

- Human motor movements follow rules of dynamical systems.  
(simplified view)
  - 1) There is a motor task.
  - 2) There is an initial state.
  - 3) The state changes over time.



| Preparation      | Ball Elevation       |  | Stability     | Release |                   | Inertia |
|------------------|----------------------|--|---------------|---------|-------------------|---------|
| Jump Preparation | Impulse for the Jump |  | Upward Flight |         | Descending Flight | Landing |

Figure 4. Phases of the jump shot.

# Dynamical system

- Human motor movements follow rules of dynamical systems.

- 1) There is a motor task.



- 2) There is an initial state.



- 3) The state changes over time.

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

“How fast and stable a movement is toward a goal”



# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

“How fast and stable a movement is toward a goal”

$k$

$b$

$x_0$

$b$

A damping coefficient shapes the **fluctuating pattern** of reaching.

$k$

A stiffness coefficient affects **rapidity** of reaching.

$x_0$

A target value indicates the **distance** between the target and the current position.

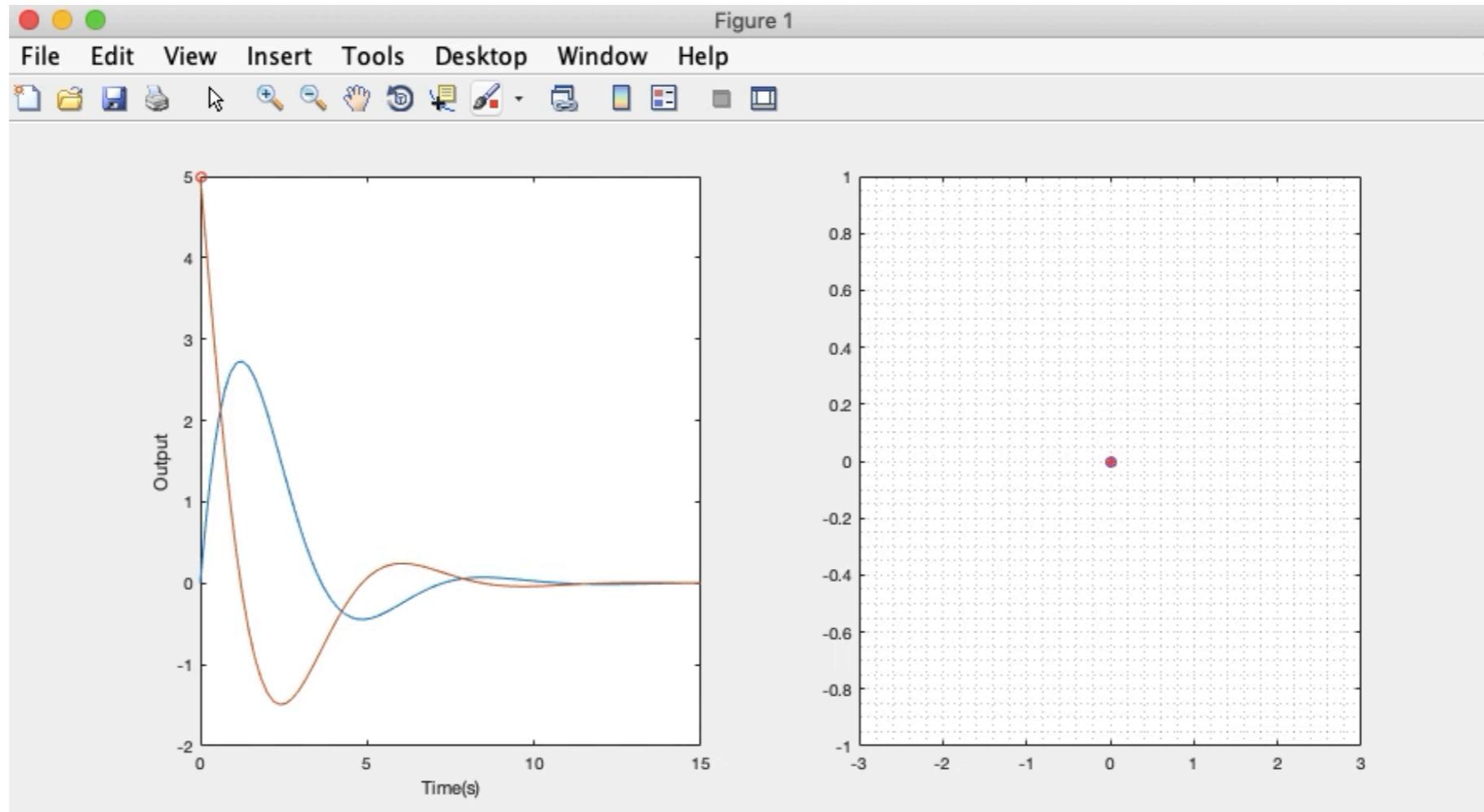
# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

$$b = 1$$

$$k = 1$$

$$x_0 = 0$$



orange: position  
blue: velocity

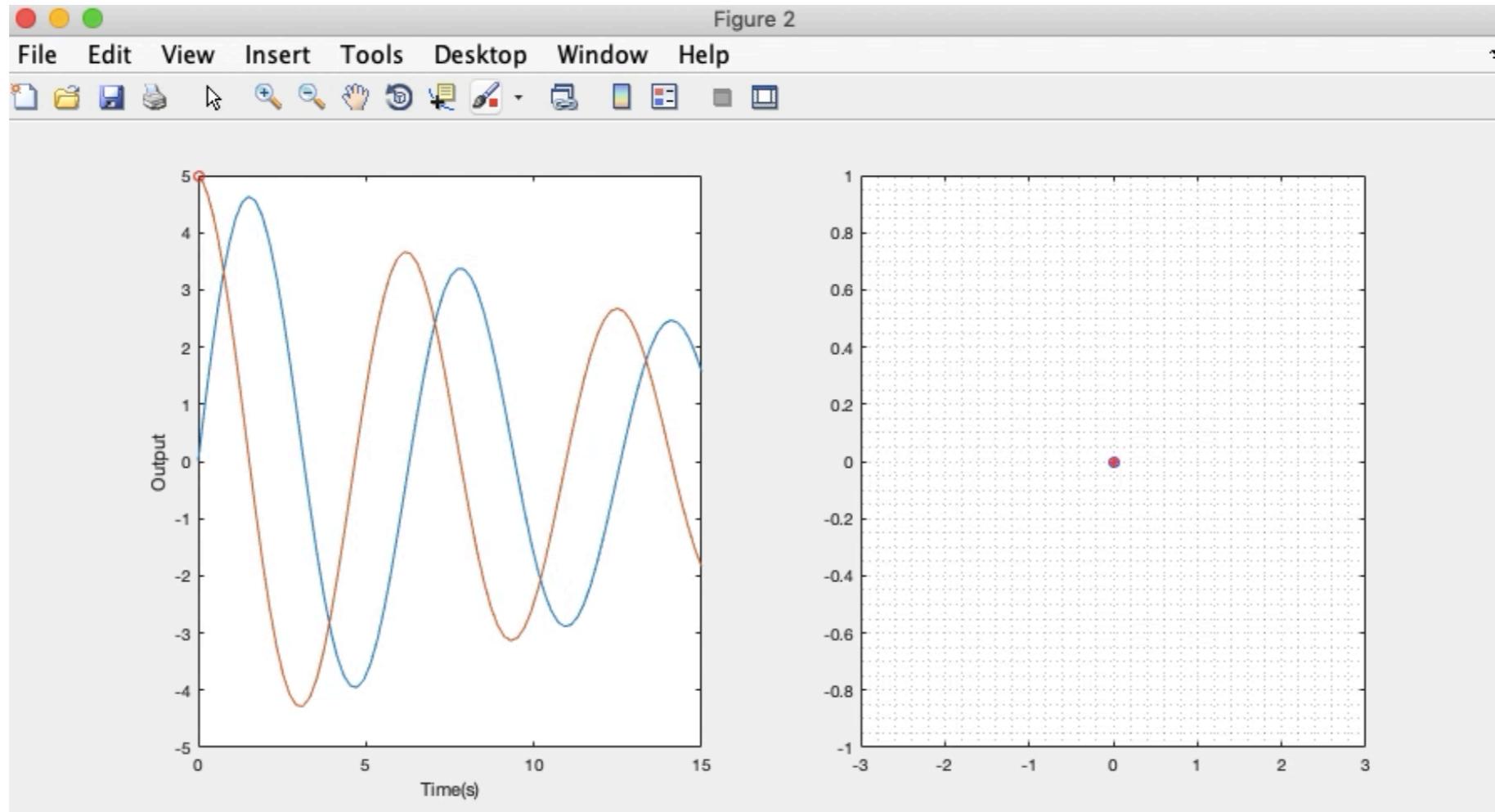
# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

$b$  = 0.1

$k$  = 1

$x_0$  = 0



orange: position  
blue: velocity

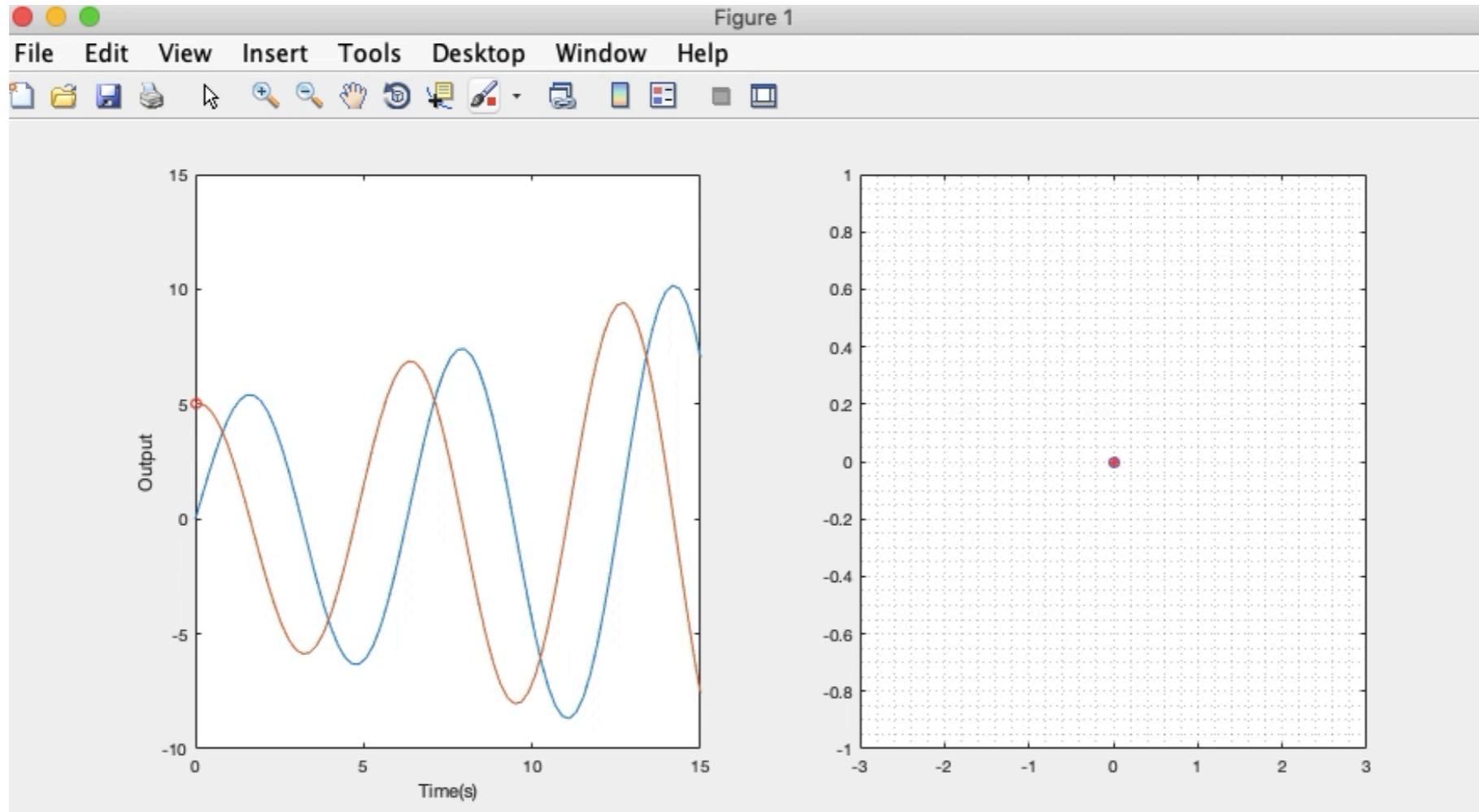
# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

$$b = -0.1$$

$$k = 1$$

$$x_0 = 0$$



orange: position  
blue: velocity

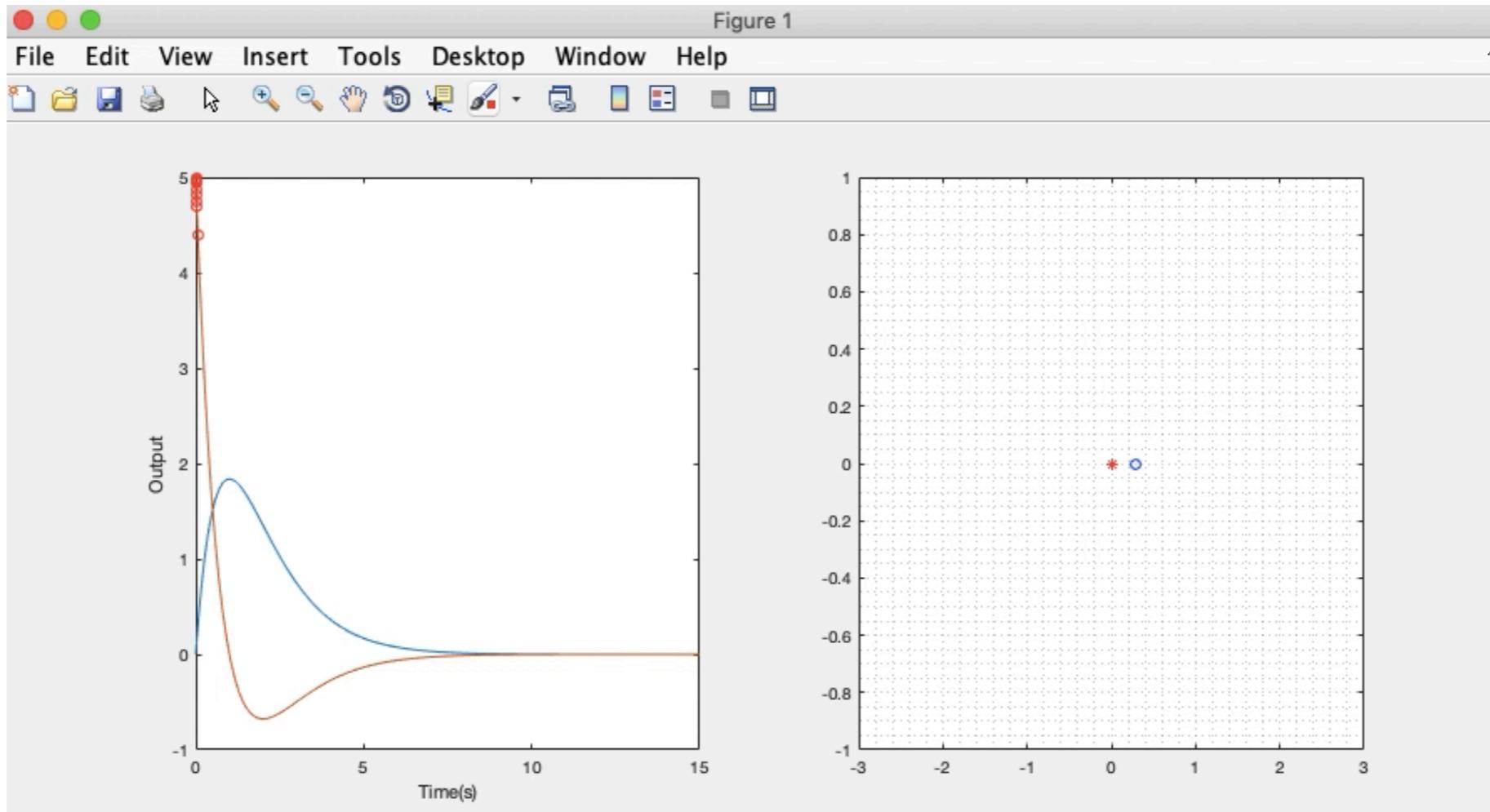
# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

$$b = 1$$

$$k = 2$$

$$x_0 = 0$$



orange: position  
blue: velocity

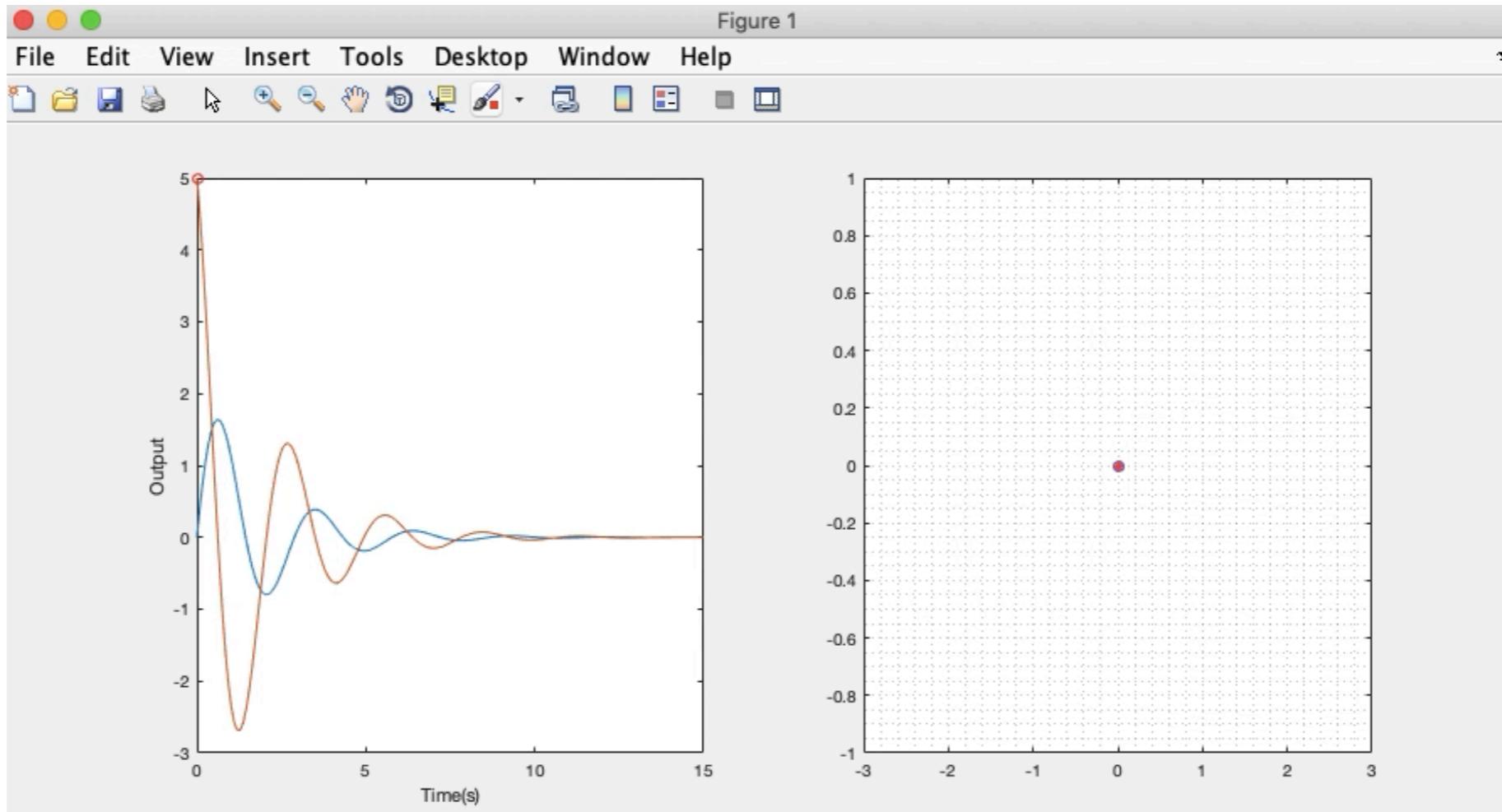
# Dynamical system

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

$$b = 1$$

$$k = 0.1$$

$$x_0 = 0$$



orange: position  
blue: velocity

# Speech production as a dynamical system

- Producing speech is also task-oriented motor movements.
- For example, each vowel or consonant has its own motor “task” to achieve a goal or vocal-tract action.
- These parameters ( $b$   $k$   $x_0$ ) are **linguistic**.
  - Invariant over lifetime of system.
  - Invariant over contexts.

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

# Speech production as a dynamical system

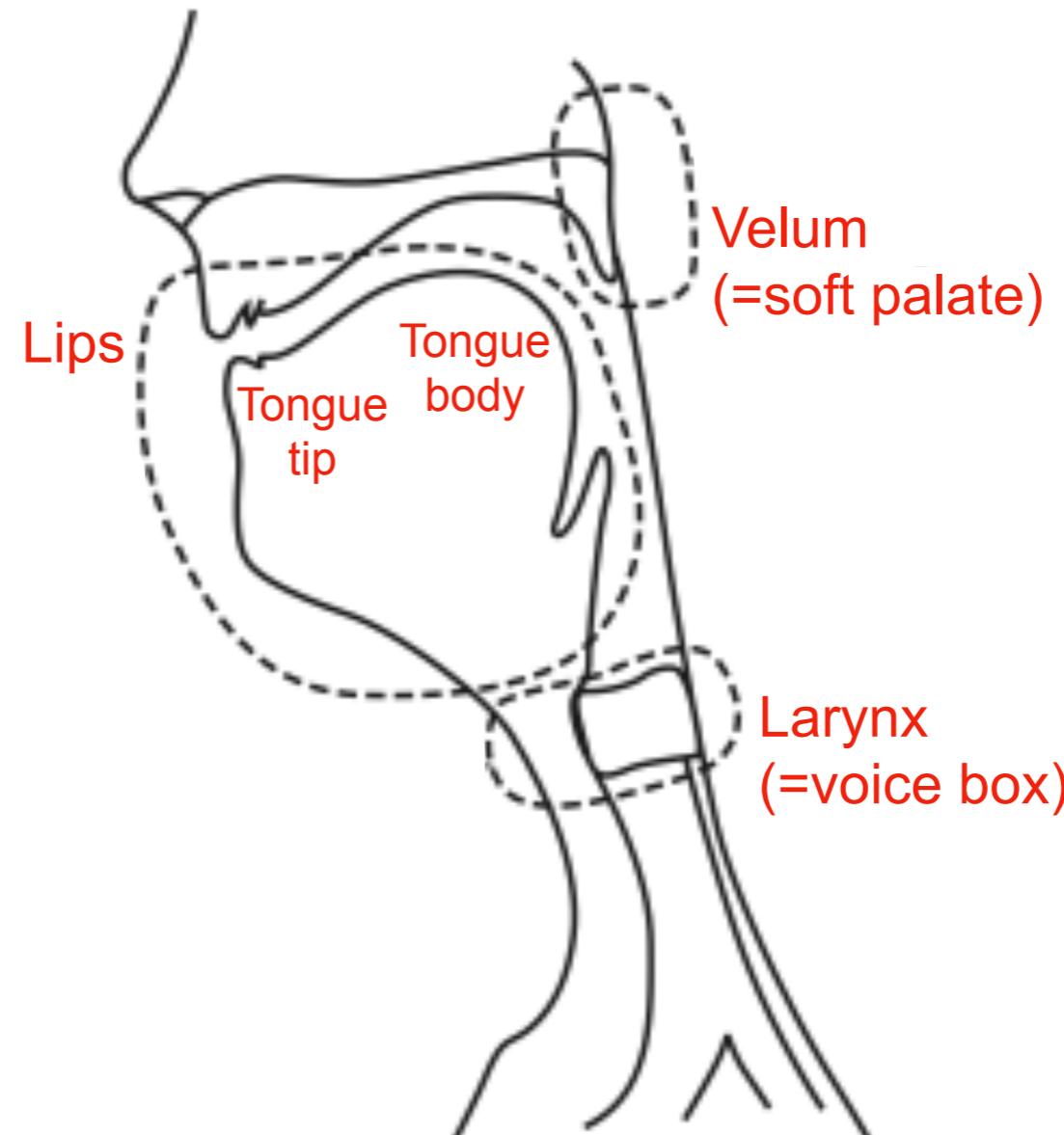
- Producing speech is also task-oriented motor movements.
- For example, each vowel or consonant has its own motor “task” to achieve a goal or vocal-tract action.
- These parameters ( $b$   $k$   $x_0$ ) are **linguistic**.
  - Invariant over lifetime of system.
  - Invariant over contexts.

$$\ddot{m} = -\frac{b}{m}\dot{x} - \frac{k}{m}(x - x_0)$$

Then, what is  $x$  and  $x_0$  in producing speech? 38

What is  $\mathcal{X}$  in producing speech?

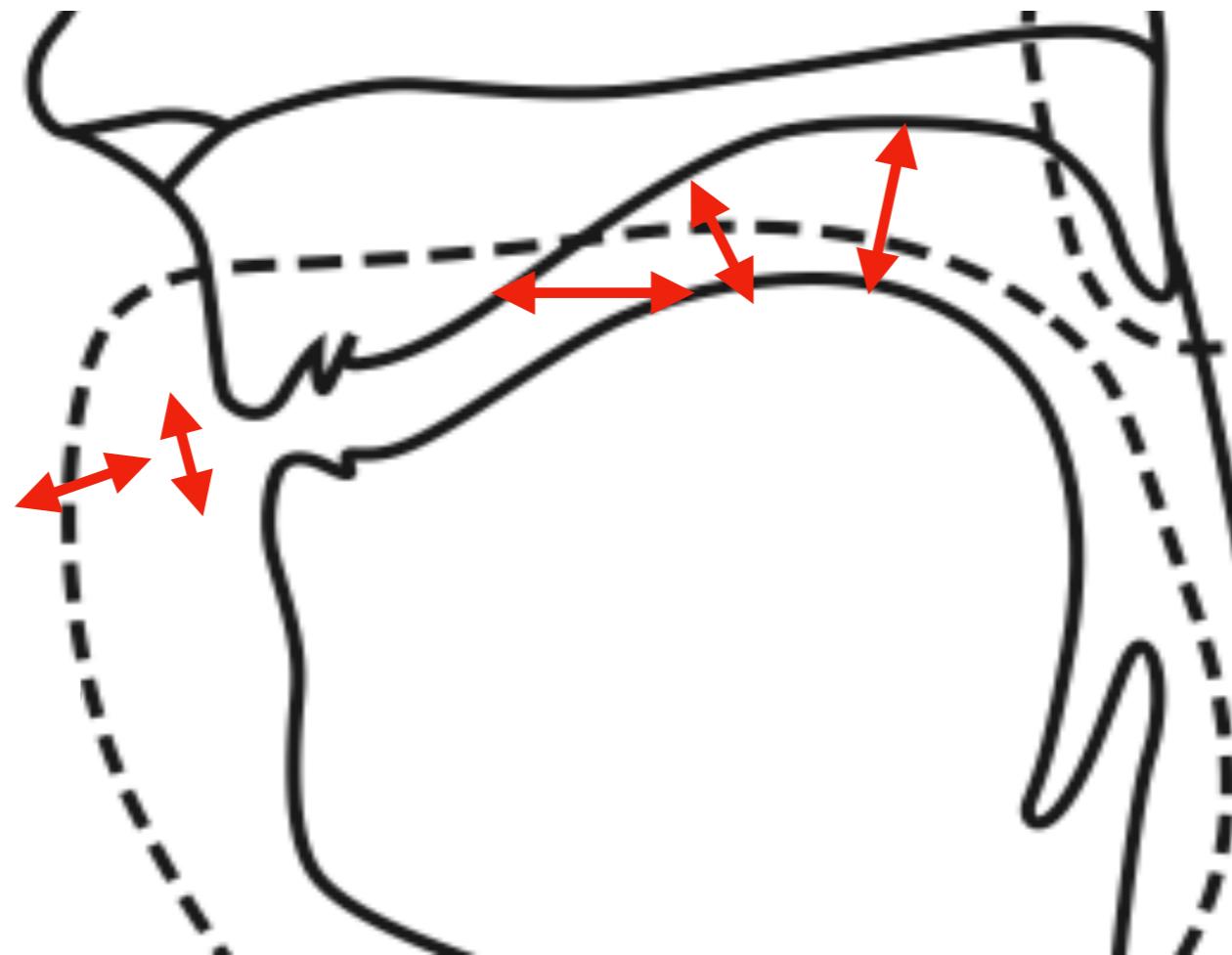
# Speech organs (=articulators)



- Five major articulators produce gestures of speech.
  - Lips, tongue tip, tongue body, velum and larynx.

What is  $x_0$  in producing speech?

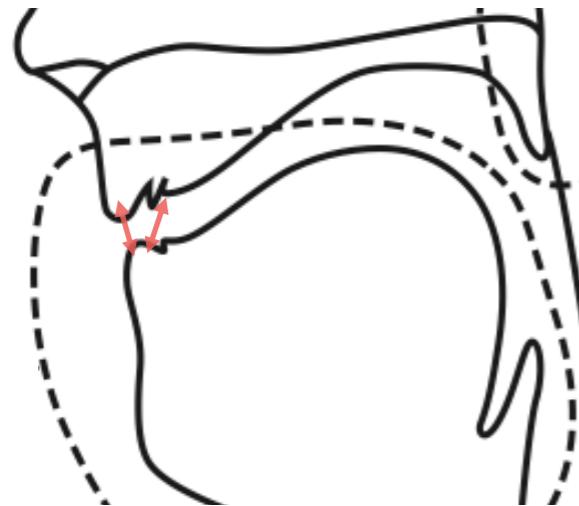
# “Task” in speech



- Making a vocal-tract constriction is a task in speech.
  - Constriction Location (CL)
  - Constriction Degree (CD)

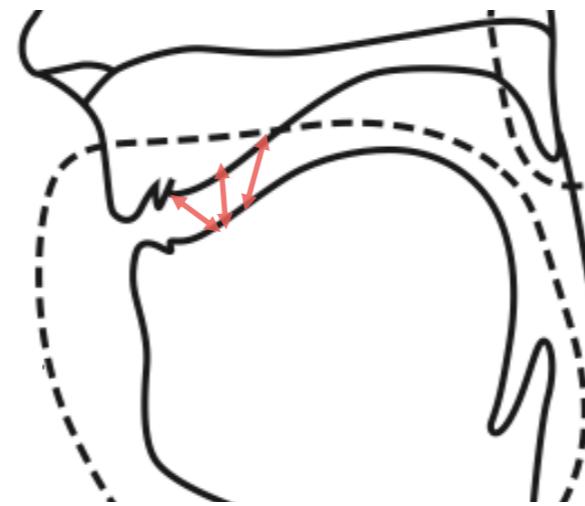
# Constriction Location (CL)

Lips



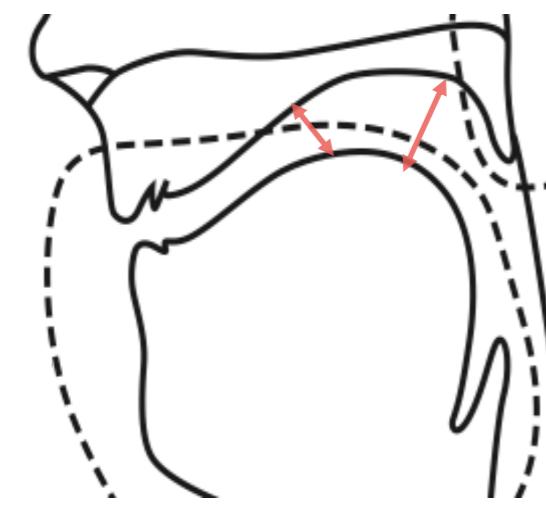
labial / labiodental

Tongue tip



dental / alveolar / palato-alveolar

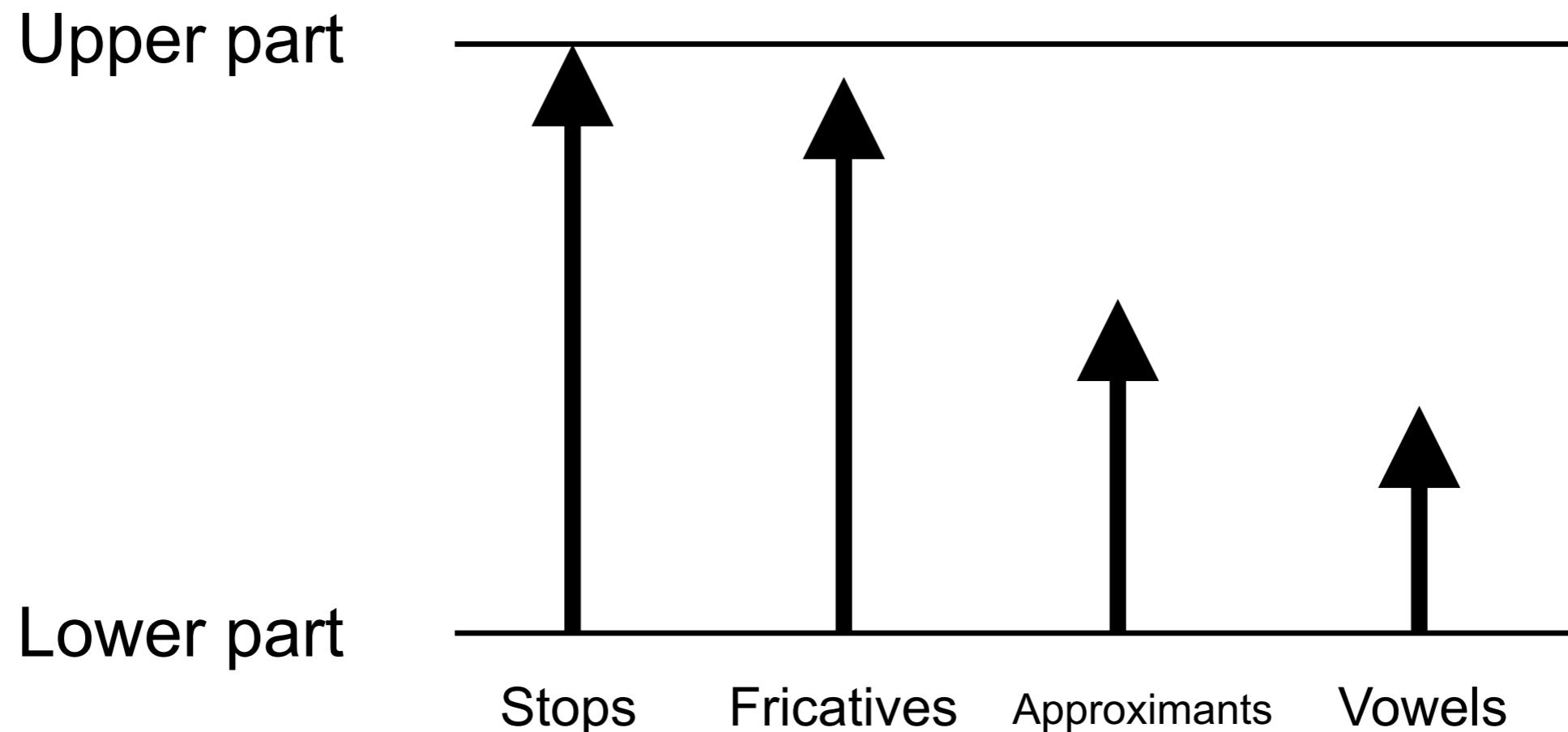
Tongue body



palatal / velar

- Constriction Location represents **the place of articulation**.

# Constriction Degree (CD)



- Constriction Degree represents **the manner of articulation**.

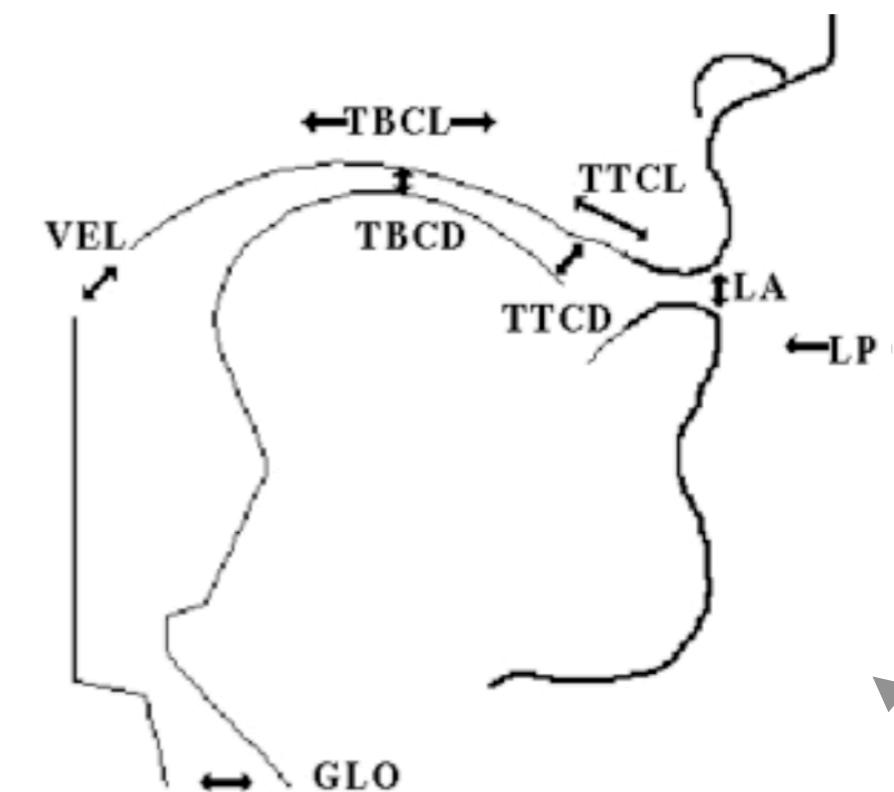
# Tasks

| Segments        | Task Variables                         | Articulators                     |
|-----------------|----------------------------------------|----------------------------------|
| p b m           | Lip Aperture ( <b>LA</b> )             | Upper Lip<br>Lower Lip<br>Jaw    |
| t d n s z sh zh | Tongue Tip (TT):<br><b>TTCD, TTCL</b>  | Tongue Tip<br>Tongue Body<br>Jaw |
| k g ñ<br>vowels | Tongue Body (TB):<br><b>TBCD, TBCL</b> | Tongue Body<br>Jaw               |
| m n ñ           | Velic Aperture ( <b>VEL</b> )          | Velum                            |
| p t k s sh      | Glottal Aperture<br>( <b>GLO</b> )     | Glottal Width                    |
| uw uh ow ao     | Protrusion ( <b>PRO</b> )              | Lip Protrusion                   |

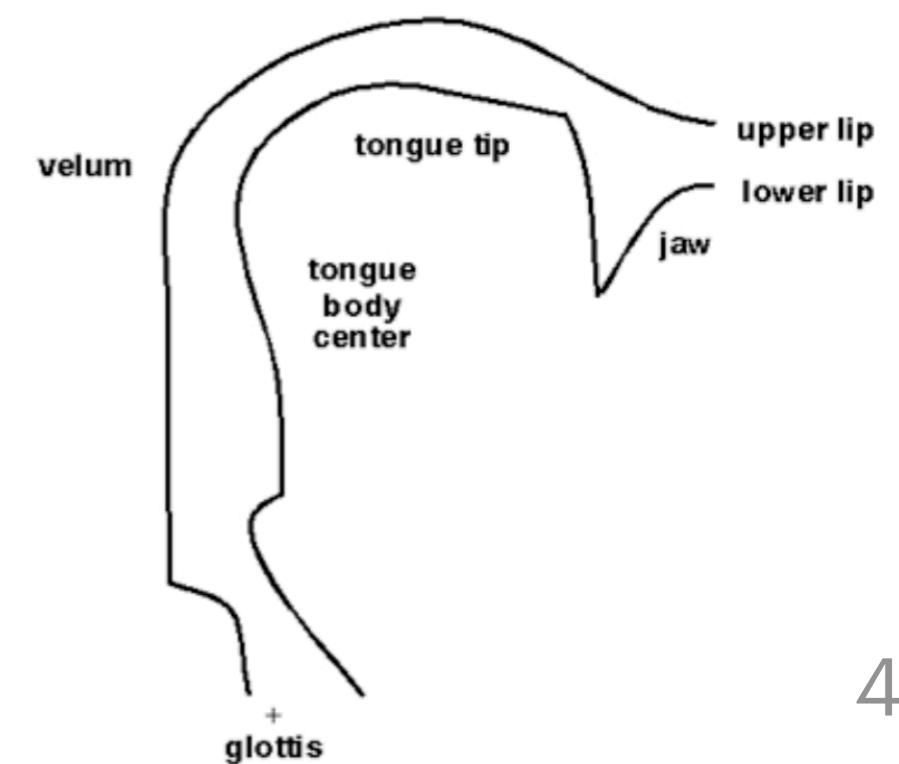
“Target”  
 $x_0$

“Input”  
 $x$

## Tasks



## Articulators

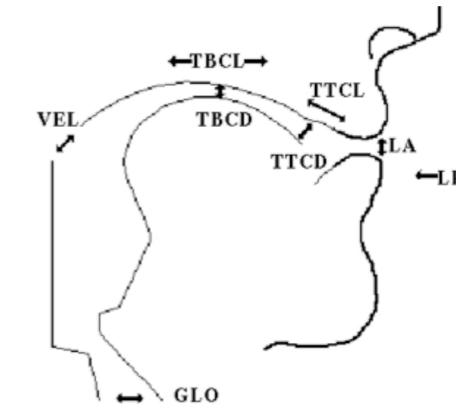


# Interim summary

- Speech gestures are time-varying vocal tract actions.
- Articulators cause constrictions at a certain location with a certain degree in the vocal tract (CD, CL).
- Movements of speech gestures follow principles of dynamical systems ( $b$   $k$   $x_0$ ).

# Exercise: /p/

Manner of Articulation

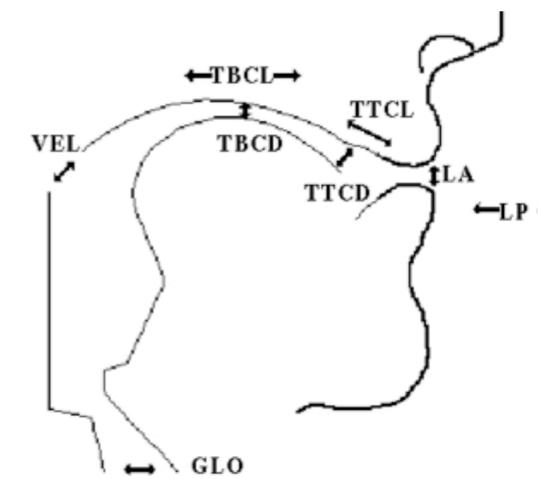


Place of Articulation

| Articulators | on/off | CD   | CL       |
|--------------|--------|------|----------|
| Lips         | on     | stop | bilabial |
| Tongue tip   | off    |      |          |
| Tongue body  | off    |      |          |
| Velum        | off    |      |          |
| Larynx       | on     |      |          |

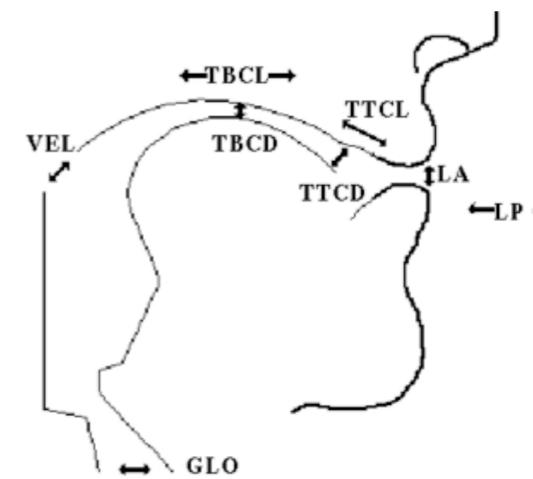
Note that the active larynx ("on") means being voiceless.

# Exercise: /t/



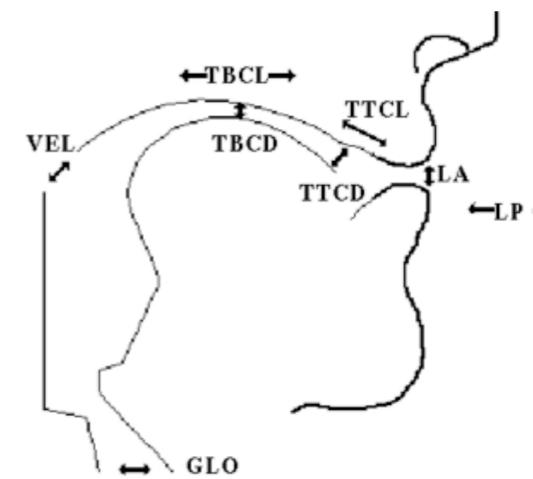
| <b>Articulators</b> | <b>on/off</b> | <b>CD</b> | <b>CL</b> |
|---------------------|---------------|-----------|-----------|
| Lips                | off           |           |           |
| Tongue tip          | on            | stop      | alveolar  |
| Tongue body         | off           |           |           |
| Velum               | off           |           |           |
| Larynx              | on            |           |           |

# Exercise: /k/



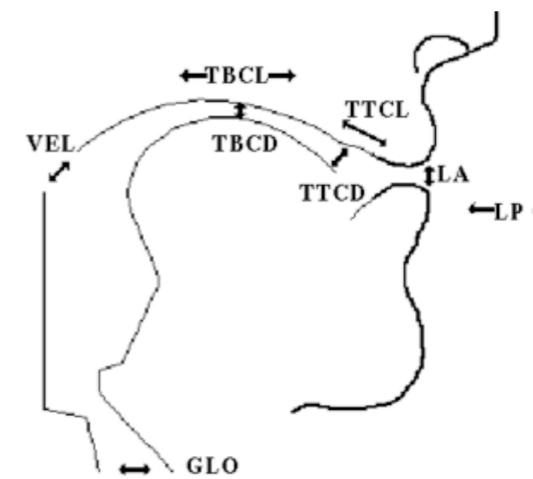
| <b>Articulators</b> | <b>on/off</b> | <b>CD</b> | <b>CL</b> |
|---------------------|---------------|-----------|-----------|
| Lips                | off           |           |           |
| Tongue tip          | off           |           |           |
| Tongue body         | on            | stop      | velar     |
| Velum               | off           |           |           |
| Larynx              | on            |           |           |

# Exercise: /g/



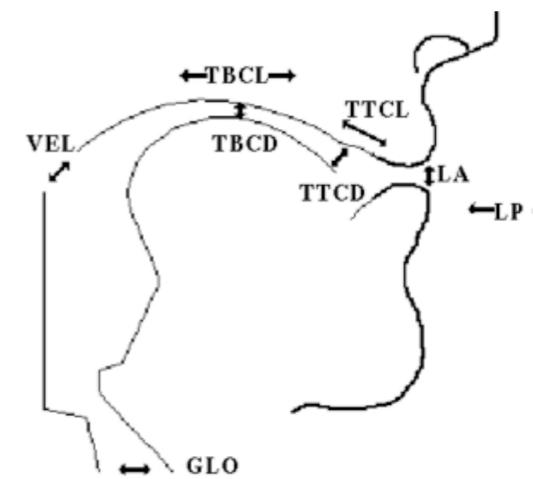
| <b>Articulators</b> | <b>on/off</b> | <b>CD</b> | <b>CL</b> |
|---------------------|---------------|-----------|-----------|
| Lips                |               |           |           |
| Tongue tip          |               |           |           |
| Tongue body         |               |           |           |
| Velum               |               |           |           |
| Larynx              |               |           |           |

# Exercise: /n/



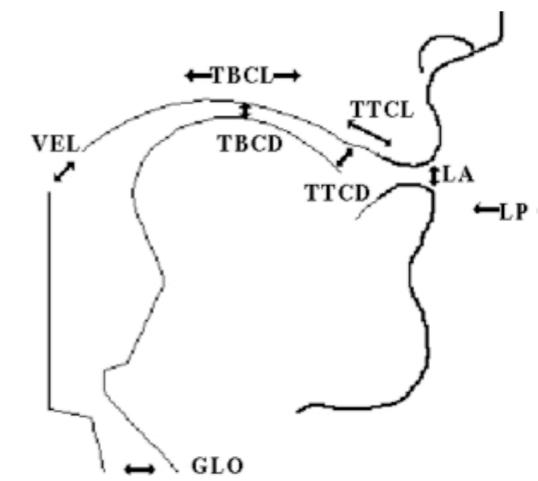
| Articulators | on/off | CD | CL |
|--------------|--------|----|----|
| Lips         |        |    |    |
| Tongue tip   |        |    |    |
| Tongue body  |        |    |    |
| Velum        |        |    |    |
| Larynx       |        |    |    |

# Exercise: /h/



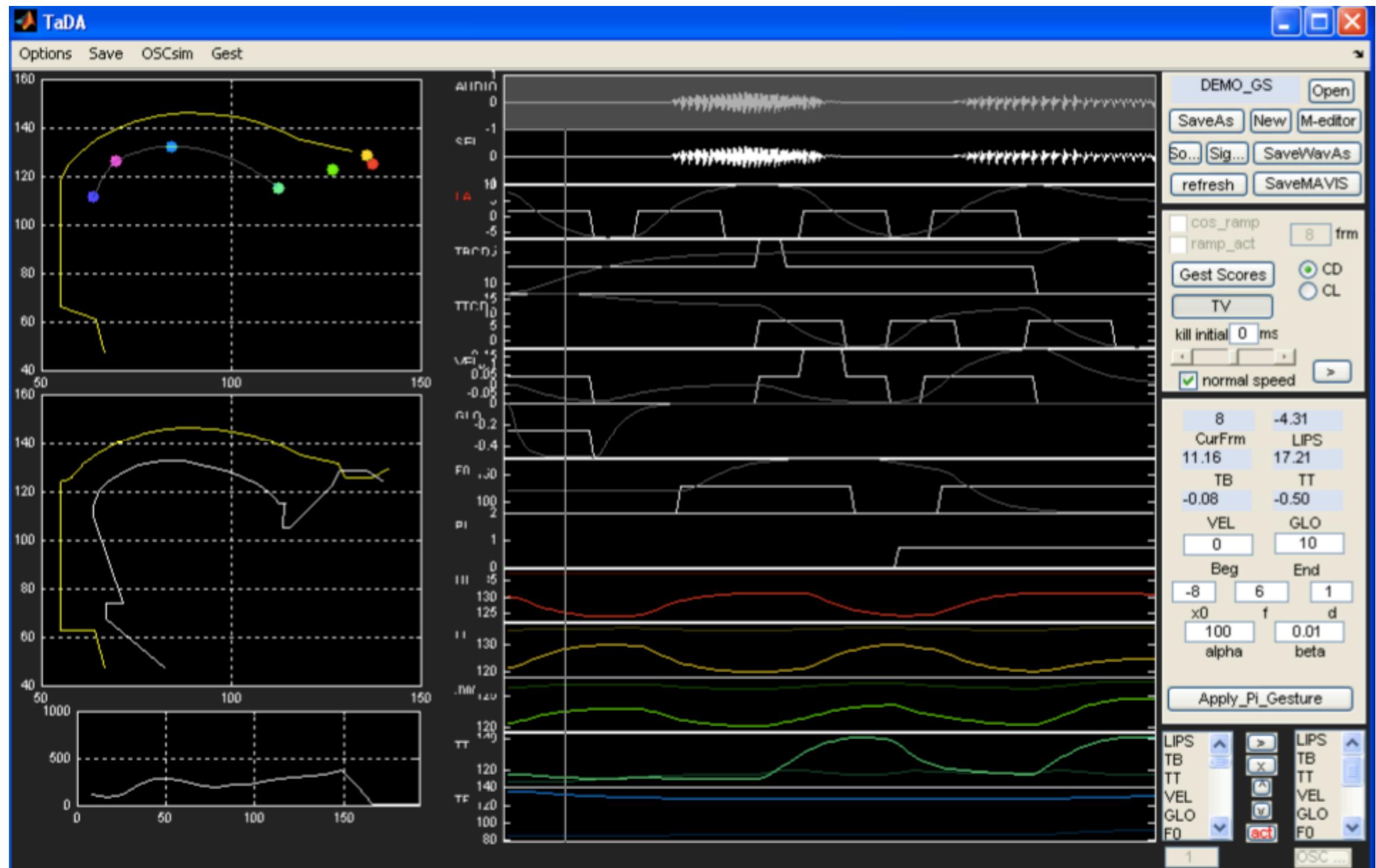
| Articulators | on/off | CD | CL |
|--------------|--------|----|----|
| Lips         |        |    |    |
| Tongue tip   |        |    |    |
| Tongue body  |        |    |    |
| Velum        |        |    |    |
| Larynx       |        |    |    |

# Exercise: /w/

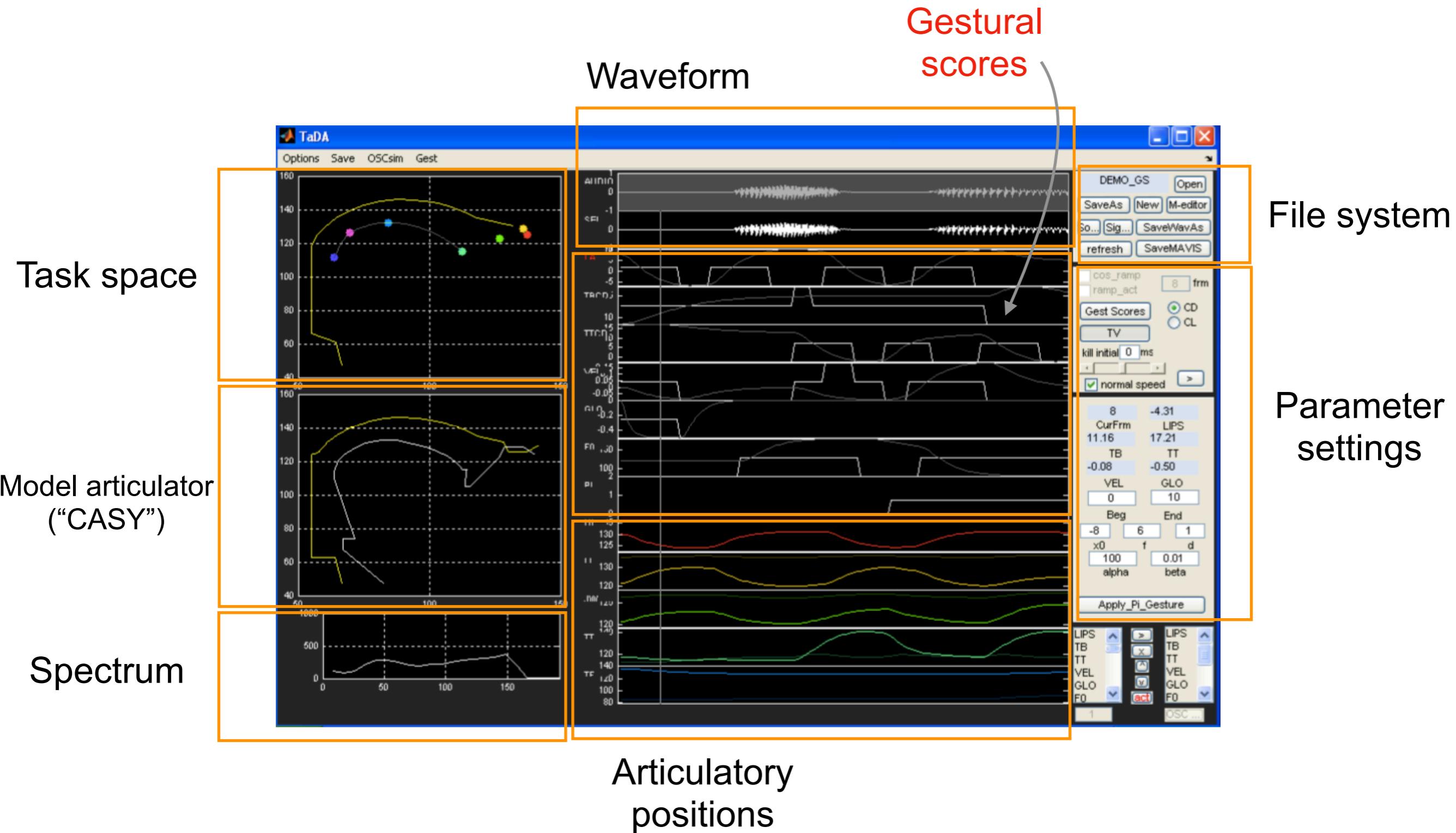


| Articulators | on/off | CD | CL |
|--------------|--------|----|----|
| Lips         |        |    |    |
| Tongue tip   |        |    |    |
| Tongue body  |        |    |    |
| Velum        |        |    |    |
| Larynx       |        |    |    |

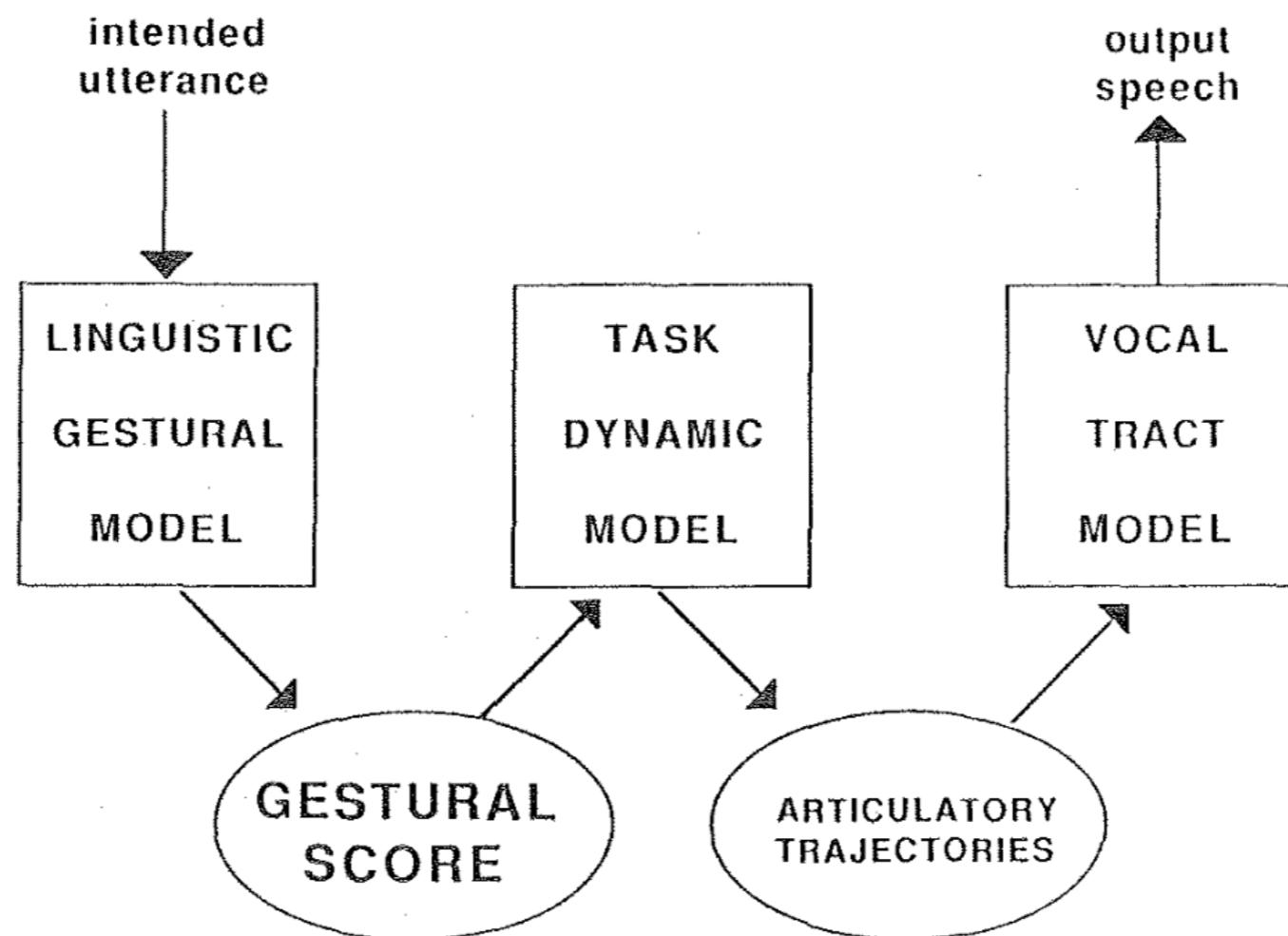
# TADA



# TADA



# Structure of TADA

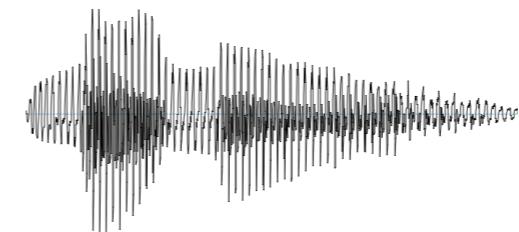


**Figure 1.** Computational modeling of gestures using articulatory dynamics.

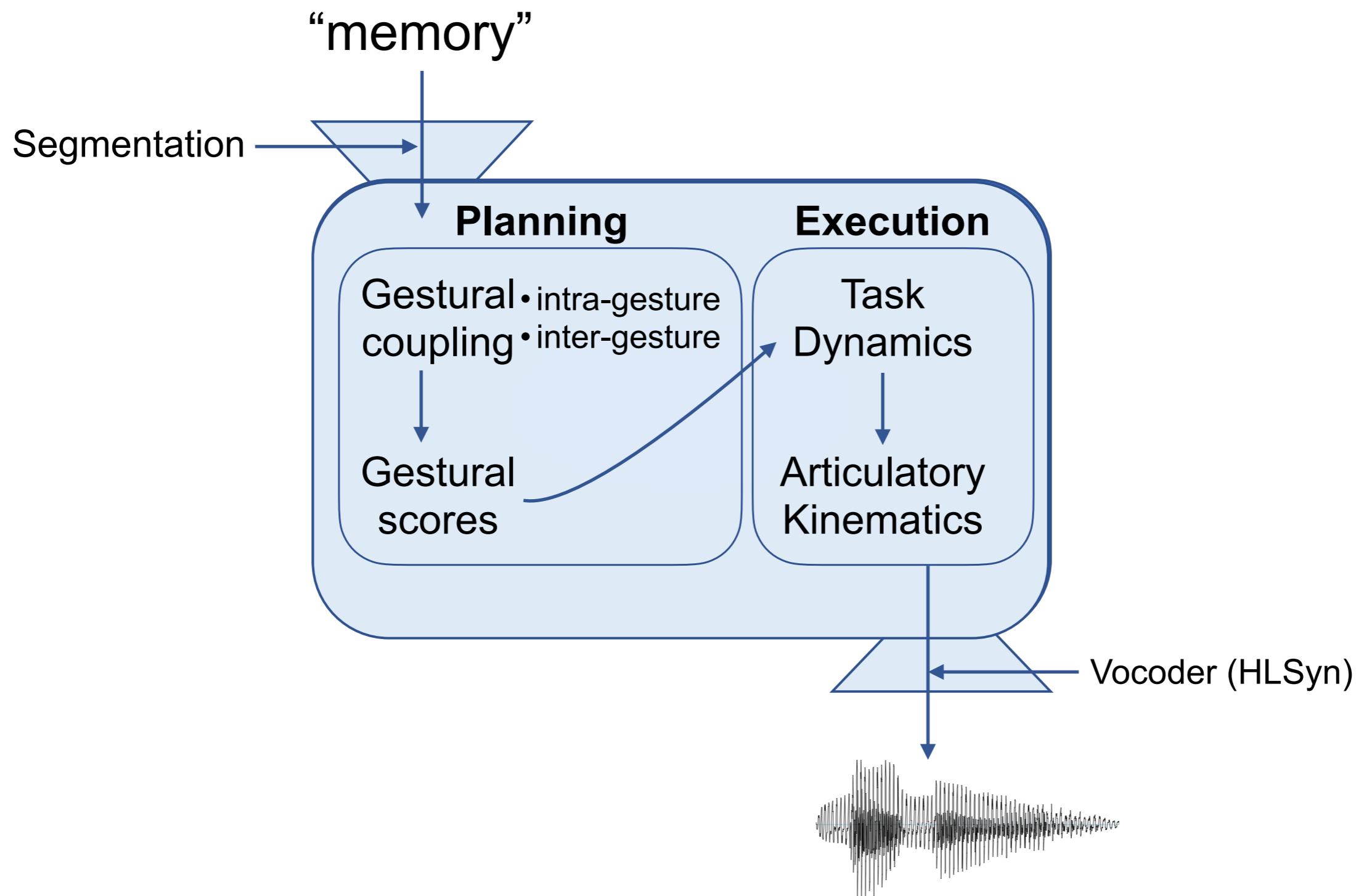
Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: between the grammar and physics of speech* (pp. 341–376). Cambridge University Press.

# Structure of TADA

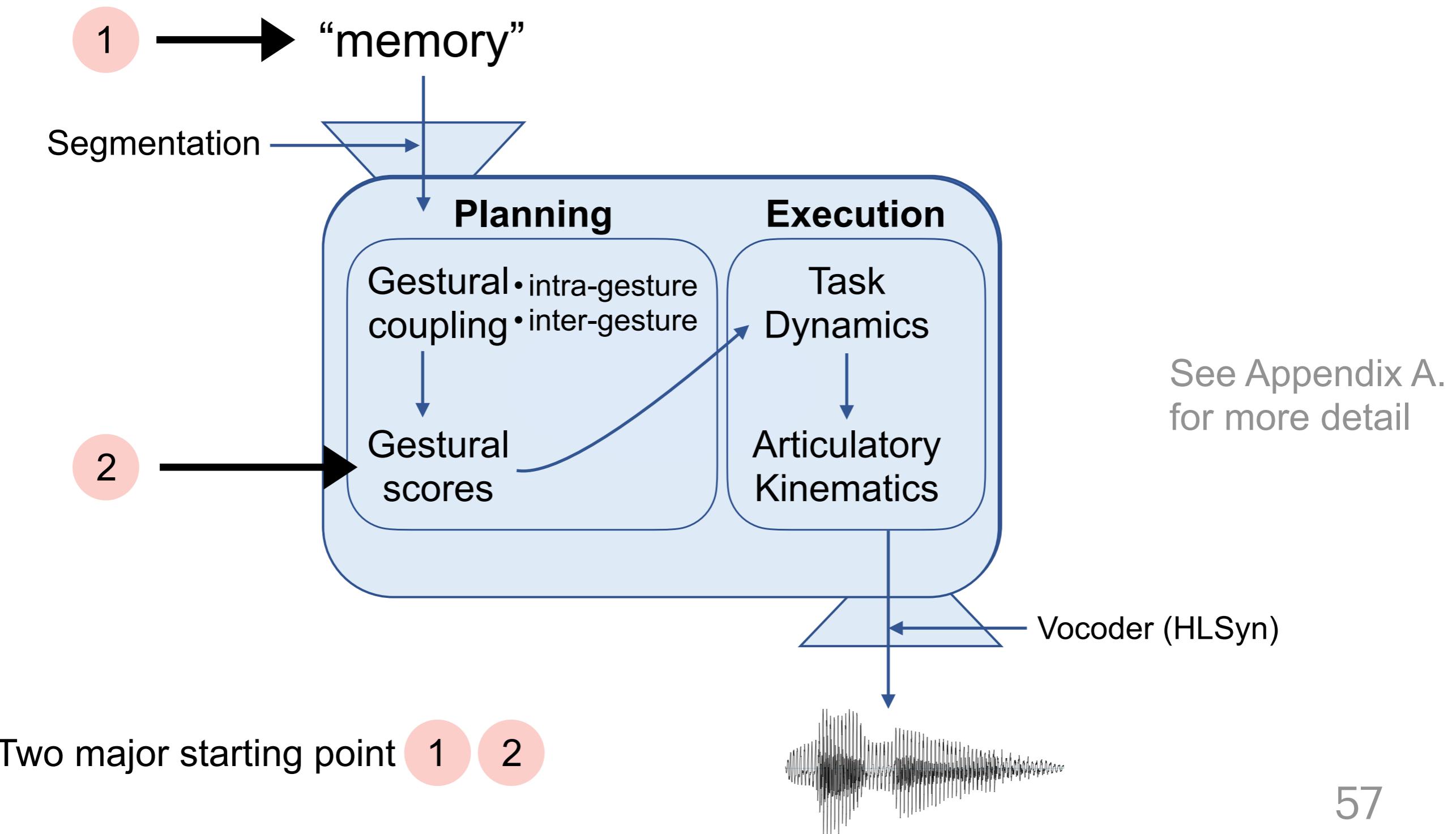
“memory”



# Structure of TADA



# Structure of TADA



# Functionalities

- Add/edit/remove gestures.
- Modify the inter-gestural and intra-gestural level of segments.
- Visually inspect articulatory configurations.
- Make non-word sounds.
- Generate the outcome (wav) file.
- Save and visualize articulatory trajectories.

# Steps for creating a single word sound

- Step 0. Open Matlab and cd to tada0.983 directory.
- Step 1. Create \*.O file by specifying a word: *Gest* —> *Run*.
- Step 2. Open \*.O file (coupling graph file).
- Step 3. Press TV button.
- Step 4. Play the sound.

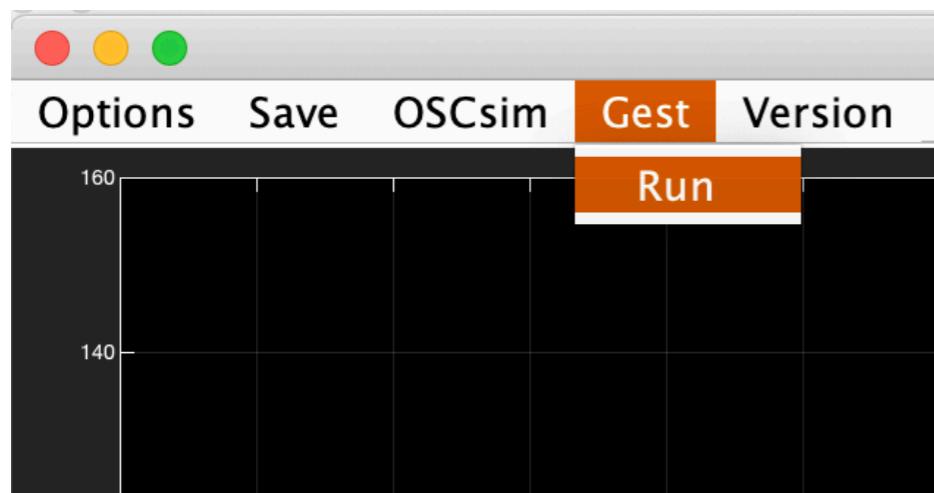
# Step 0

- Download and unzip **tada0.983.zip**.
- Open Matlab
- Change the directory to the unzipped folder ('tada0.983').
  - using cd command or
  - dragging the folder icon to the command window.
- type 'tada' in the command window.
- TADA window will open in a second.

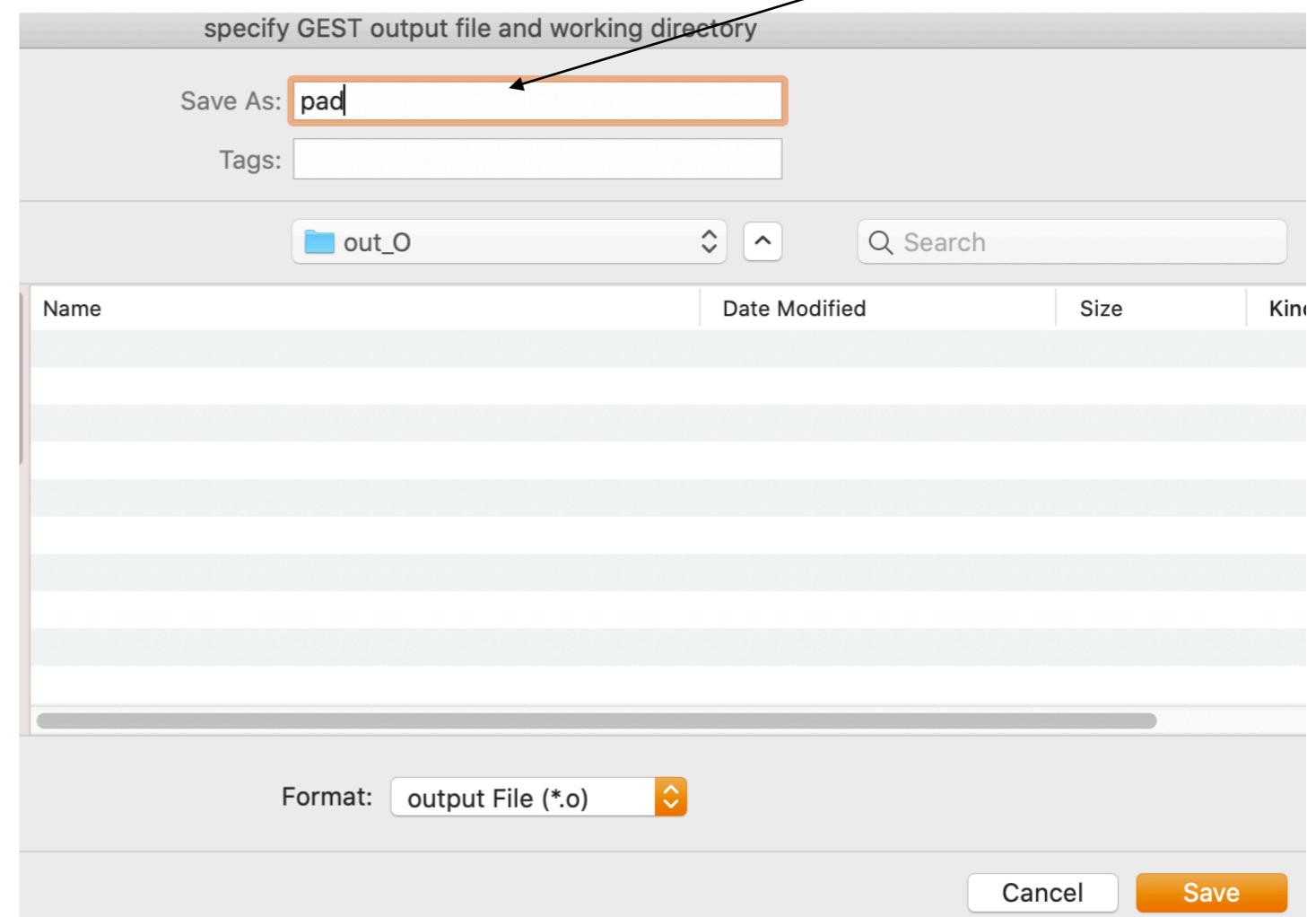
# Step 1

Synthesis of a word “pad”

- Create \*.O file by specifying a word: *Gest* → *Run*.  
Type the word



*Gest* is located on the top left corner of the TADA window.



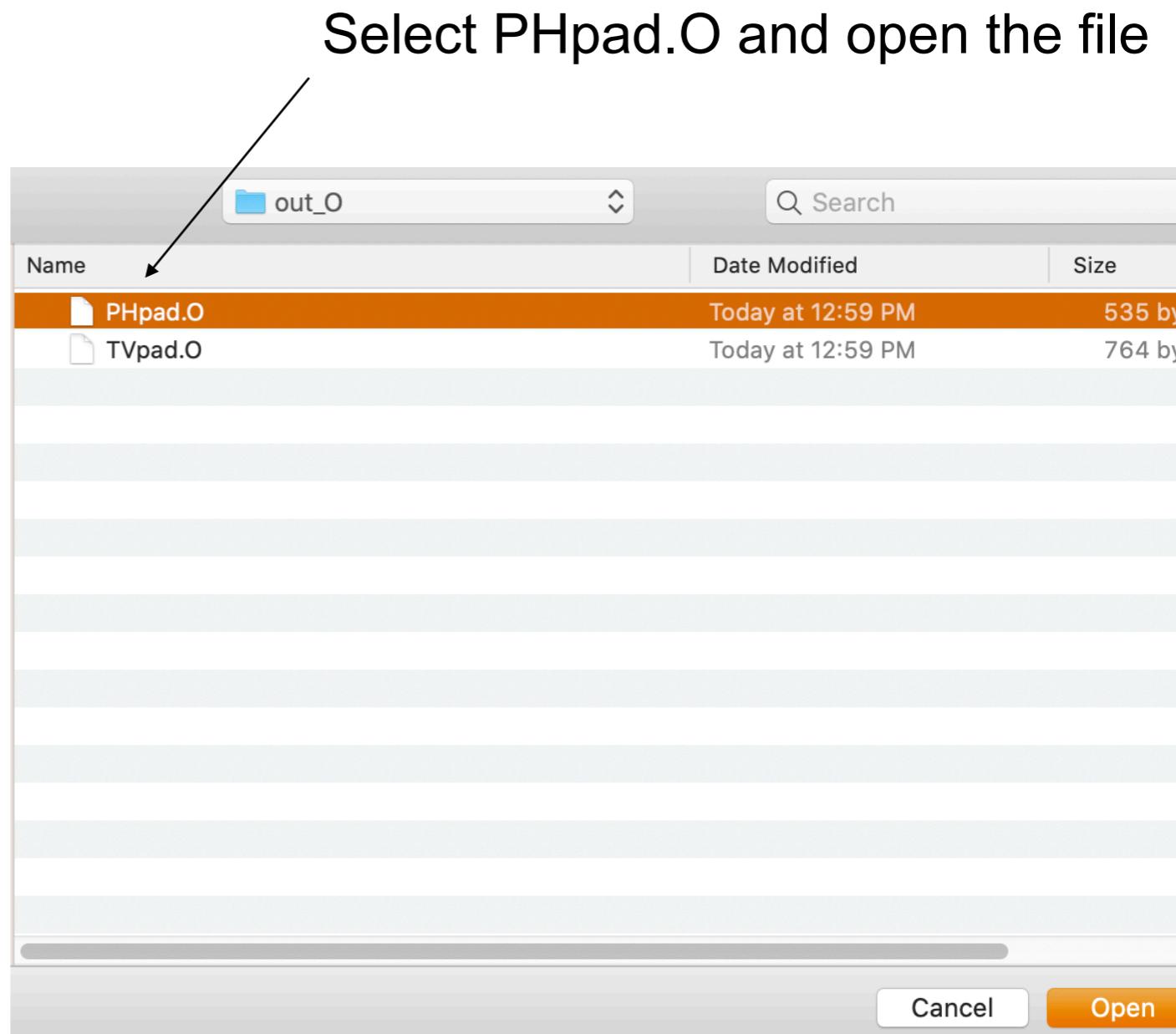
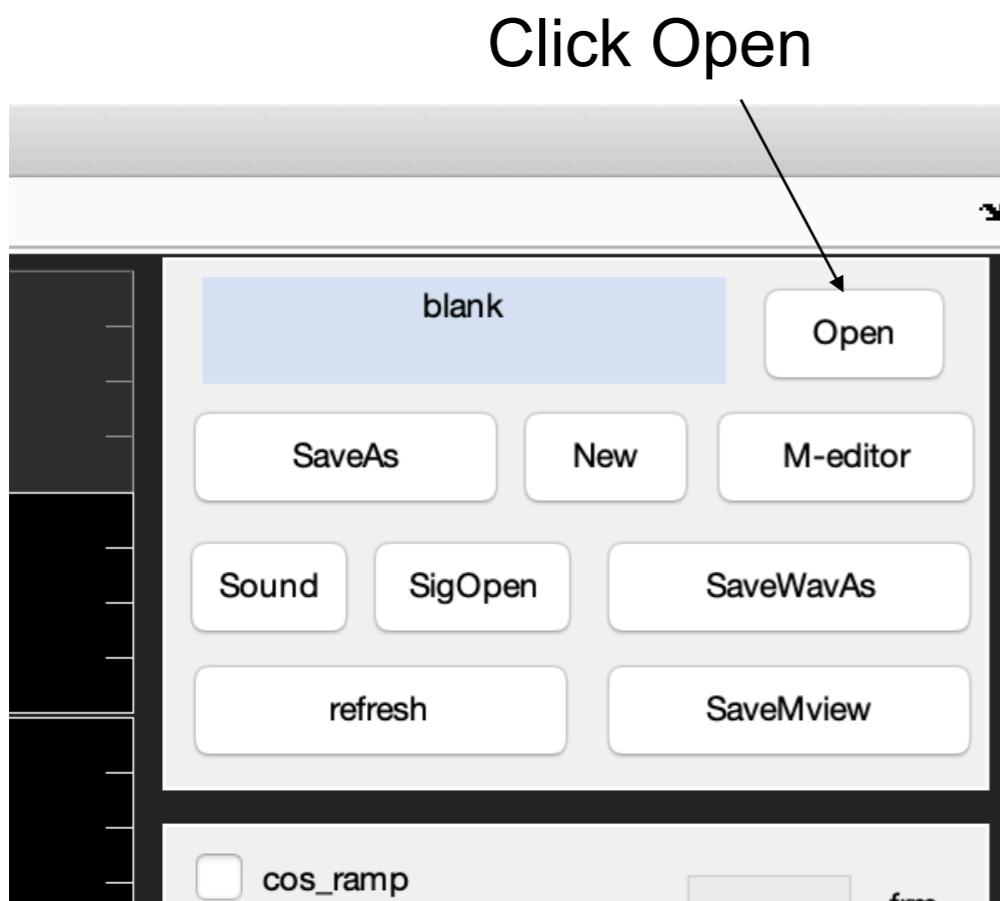
After saving, you will see another pop-up window and you can just click Okay without changing anything.

Save

61

# Step 2

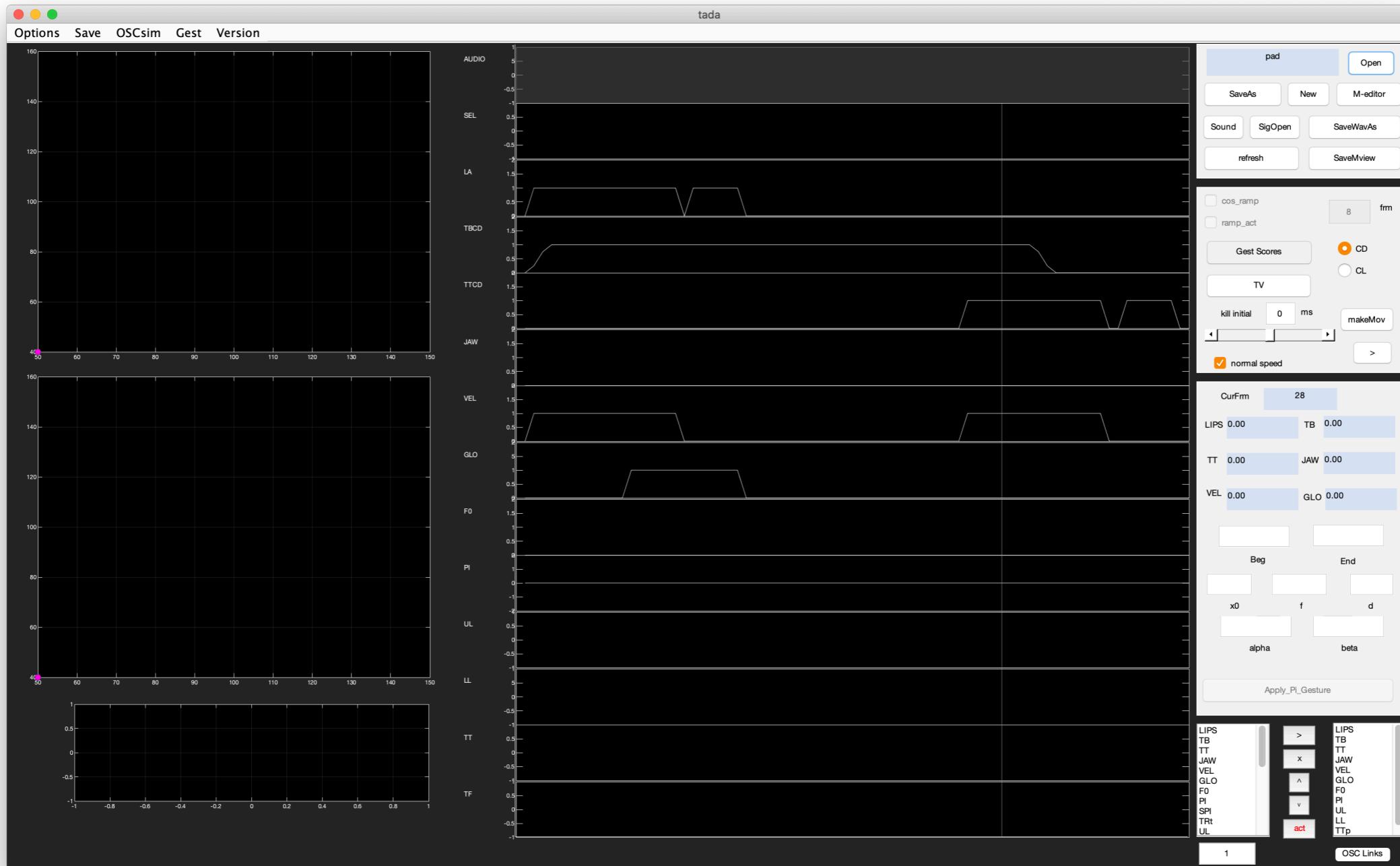
- Open \*.O file.



PHpad.O file includes information about intra-gestural relationship  
TVpad.O file includes information about inter-gestural relationship

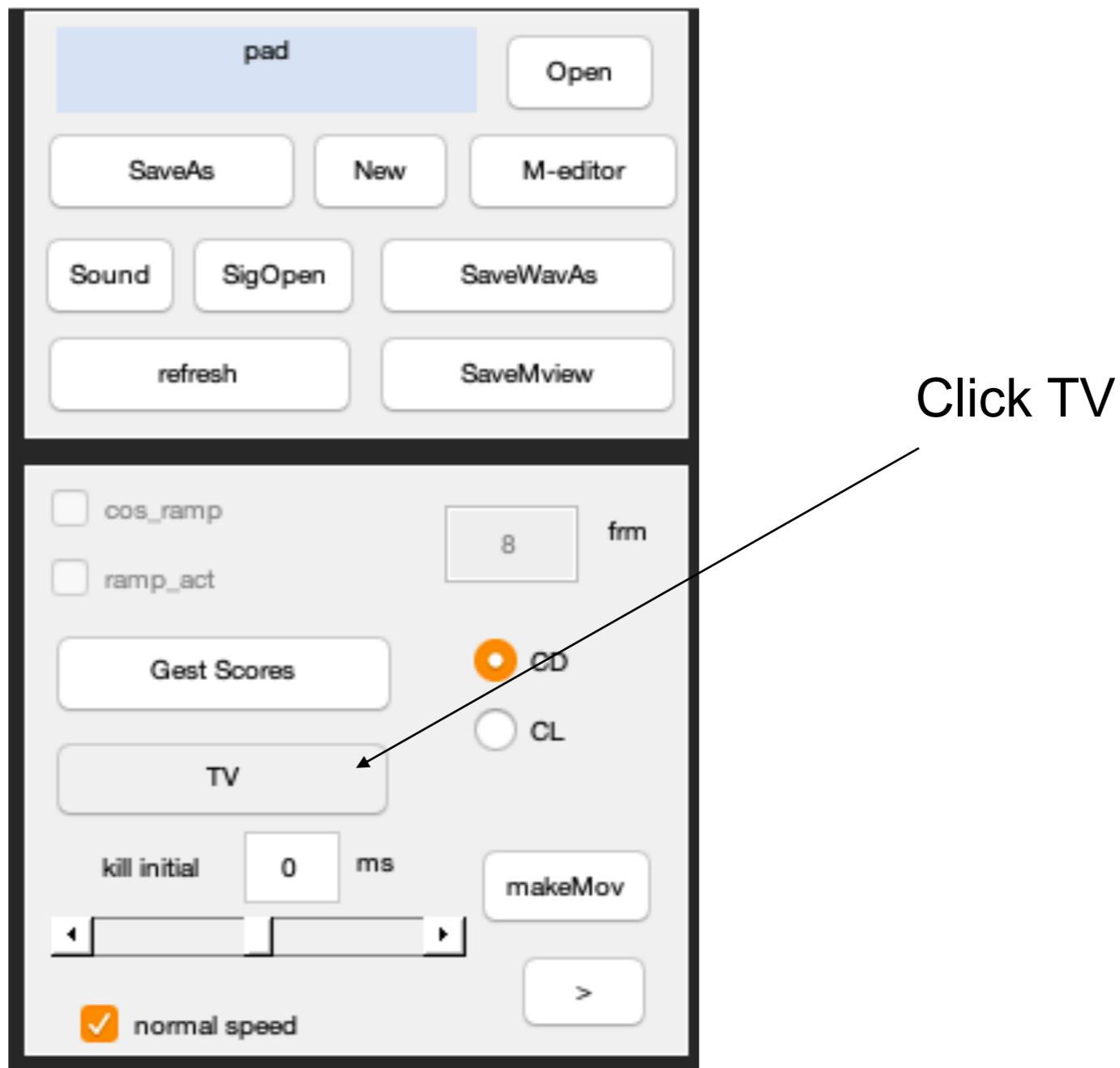
# Step 2

- After opening the \*.O file, gestural scores are loaded on the middle panel.



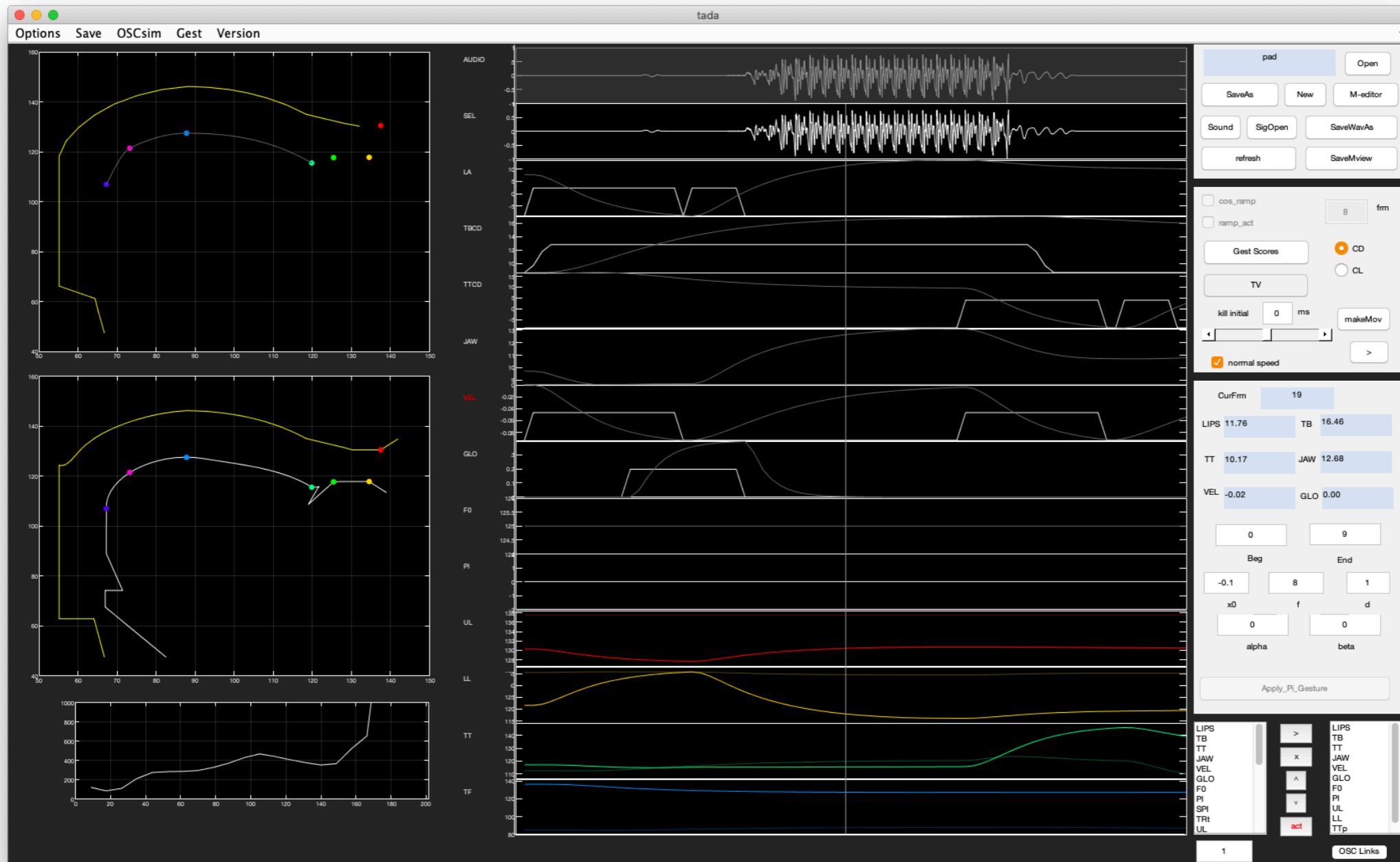
# Step 3

- Press TV button.



# Step 4

- Articulatory trajectories and the corresponding sound are created. Play with the cursor at the waveform and listen to the sound by clicking Sound button on the top right panel!



# Exercise: /pæd/, /bæd/

Fill in the blanks by specifying temporal locations of each articulator and create a sound using TADA.

**/pæd/**

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

**/bæd/**

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

# Exercise: /tæd/, /tæn/

Fill in the blanks by specifying temporal locations of each articulator and create a sound using TADA.

/tæd/

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

/tæn/

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

# Exercise: /sɪ/, /ʃɪ/

Fill in the blanks by specifying temporal locations of each articulator and create a sound using TADA.

/sɪ/

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

/ʃɪ/

|             |  |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |  |

# Exercise: *had, hid*

Fill in the blanks by specifying temporal locations of each articulator and create a sound using TADA.

***had***

|             |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |

***hid***

|             |  |  |  |  |  |  |  |  |  |  |
|-------------|--|--|--|--|--|--|--|--|--|--|
|             |  |  |  |  |  |  |  |  |  |  |
| Velum       |  |  |  |  |  |  |  |  |  |  |
| Tongue Tip  |  |  |  |  |  |  |  |  |  |  |
| Tongue Body |  |  |  |  |  |  |  |  |  |  |
| Lips        |  |  |  |  |  |  |  |  |  |  |
| Glottis     |  |  |  |  |  |  |  |  |  |  |

# Summary

- Speech production is a dynamical system where the time-varying articulatory gestures are coordinated for the constriction tasks in the vocal tract.
- Different constriction tasks can be specified by varying their location and degree.
- TADA provides ways to simulate such processes and to evaluate them through the articulatory trajectories and synthesized sound.

# For the next week

- Download TADA in the following link: .
- Do exercises in slide 66 - 69 using TADA on your own.

# References

- Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: between the grammar and physics of speech* (pp. 341–376). Cambridge University Press.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to Language Via the Mirror Neuron System*, 215–249. <https://doi.org/10.1017/CBO9780511541599.008>
- Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America*, 115(5), 2430-2430.
- Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.

# Appendix A.

## TADA synthesis flow

