An agonist-antagonist pitch production (A2P2) model

Branislav Gerazov and Philip N. Garner

Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University of Skopje, Macedonia

Idiap Research Institute, Martigny, Switzerland

Outline

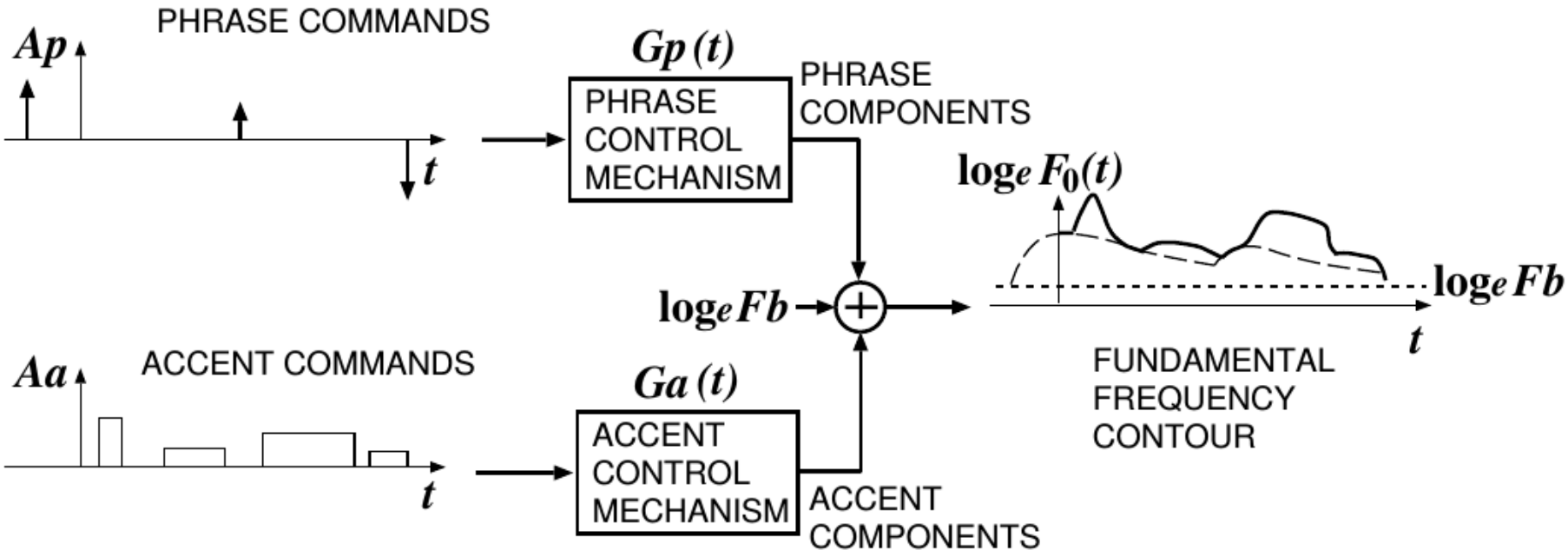
- Introduction
- SDM and Hill type models
- Agonist-antagonist pitch production model
- Conclusions

Introduction

- Prosody modelling is crucial for a number of speech technologies: TTS, SER, ASR ...
- A lot of intonation models have been proposed:
  - Surface F0 models: ToBI, INSINT, IViE, Tilt, SFC ...
  - Physiological F0 models: CR, StemML, qTA ...

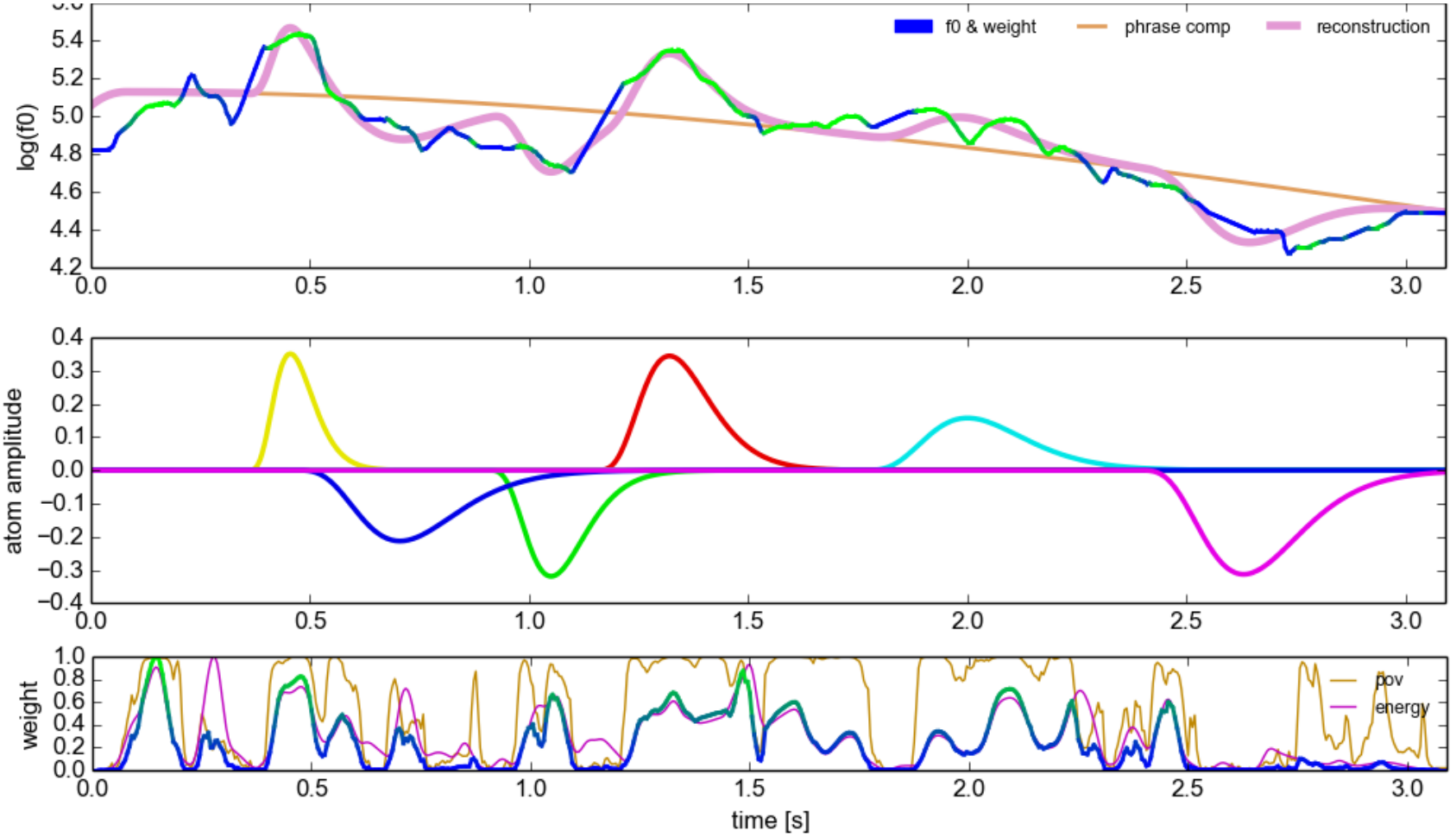
Introduction

- Fujisaki's Command Response model

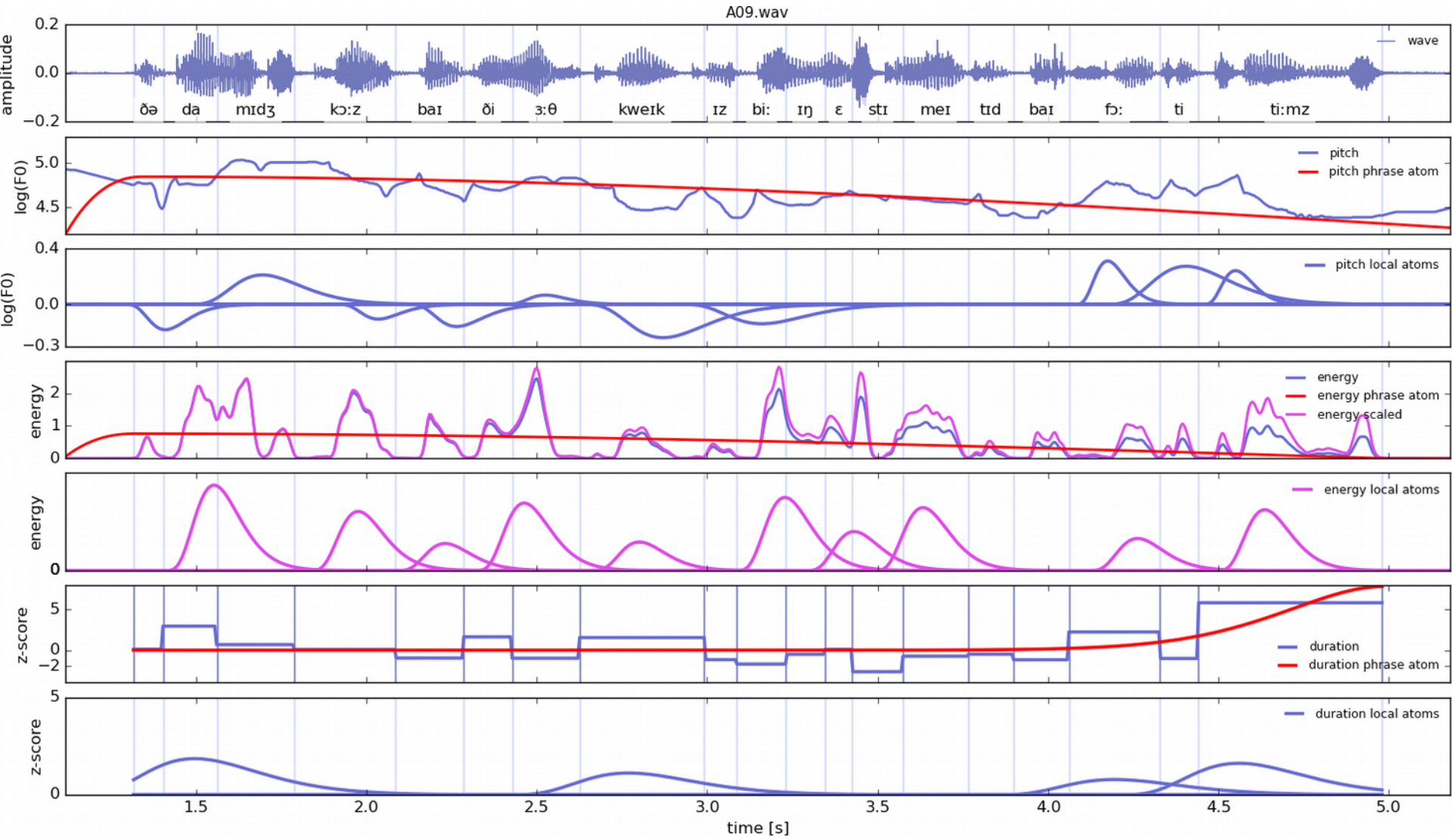


Introduction

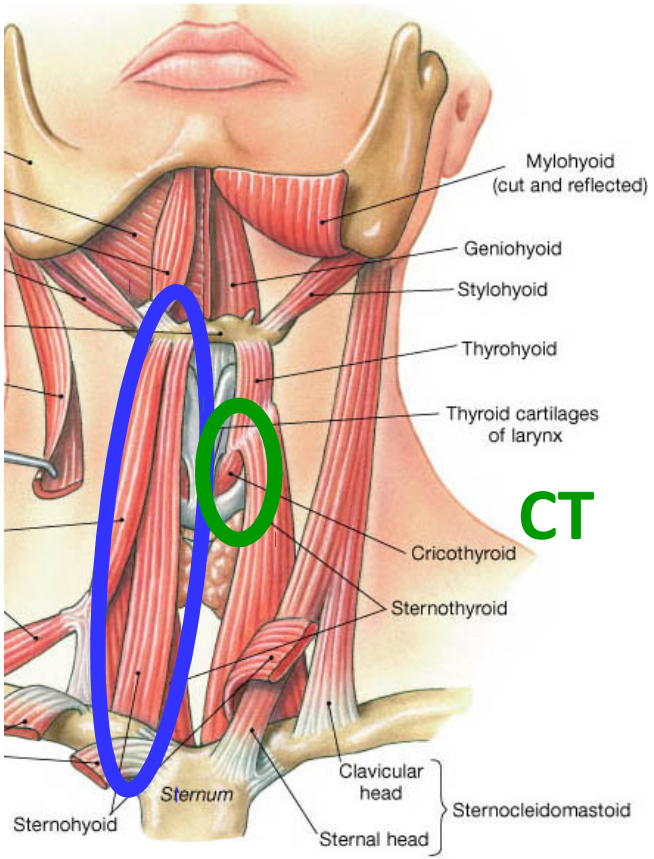
- Generalized Command Response model



Introduction

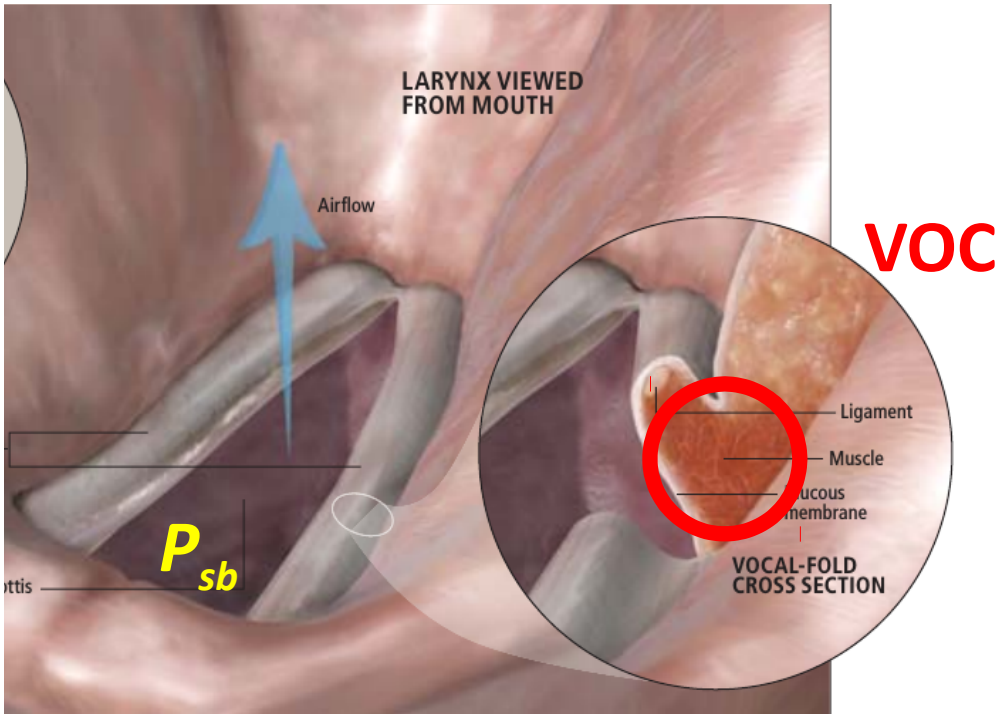


Introduction



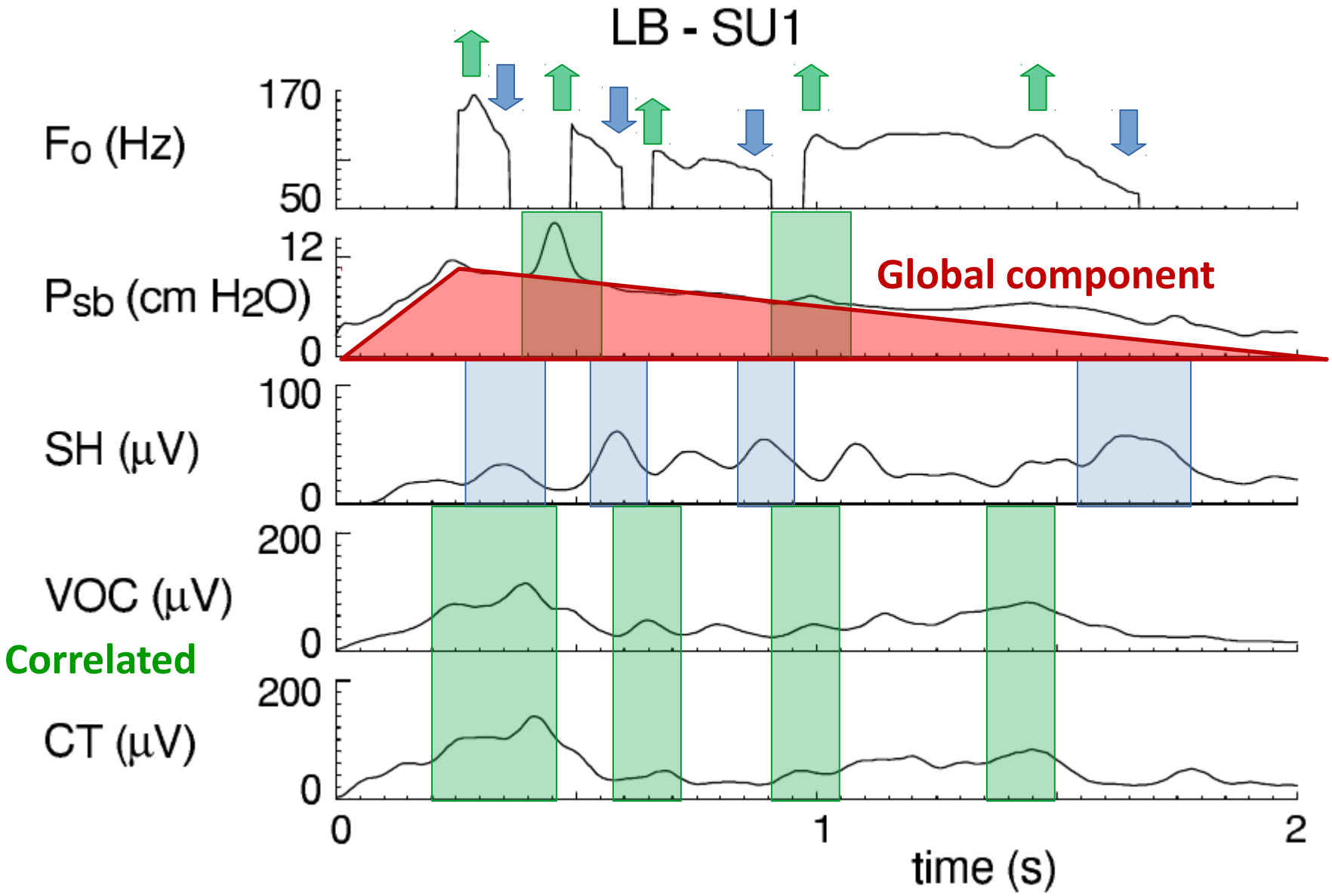
CT

SH



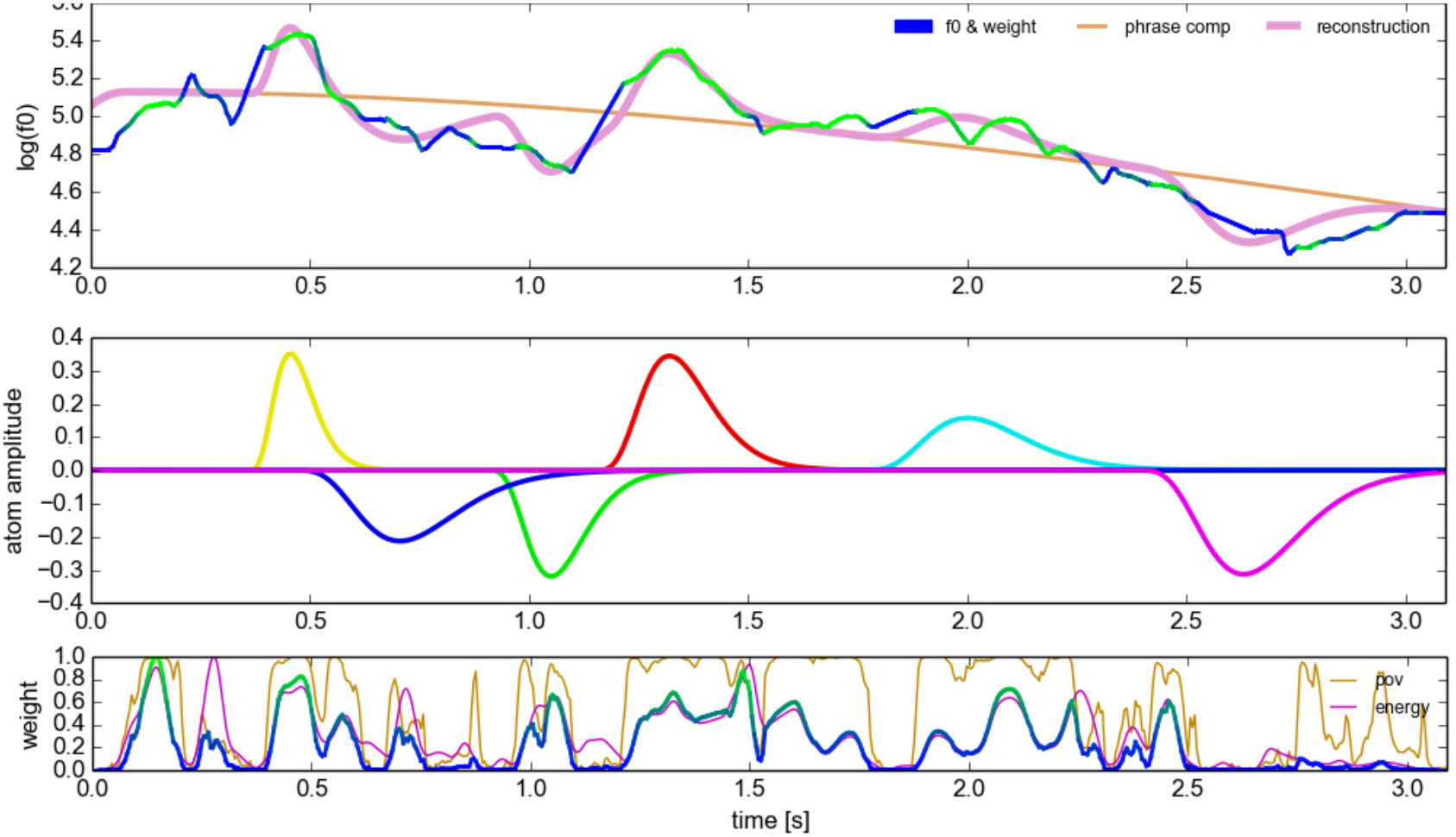
VOC

Introduction

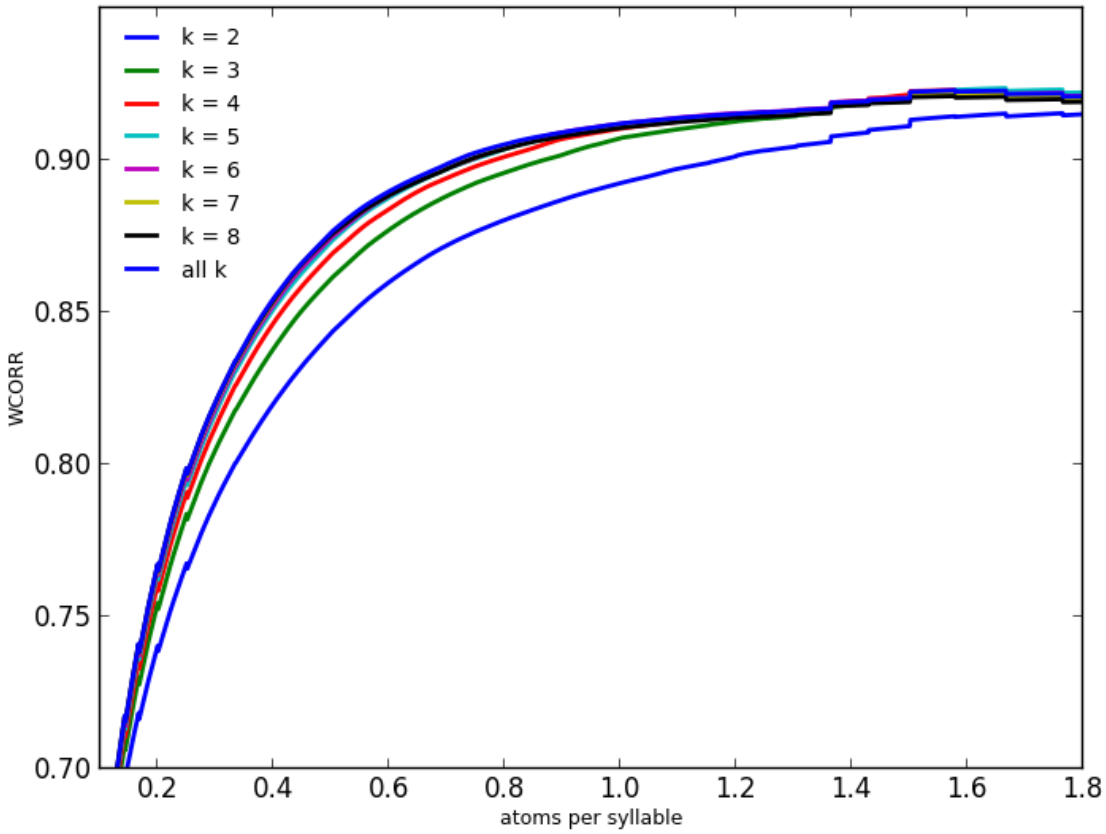


Introduction

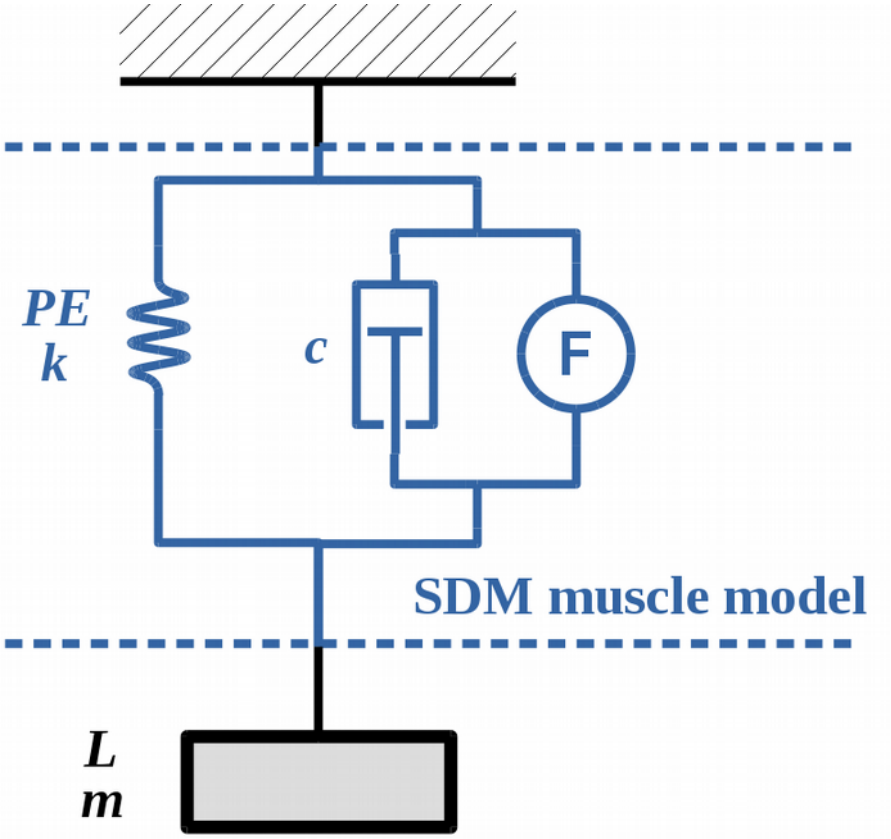
- Generalized Command Response model



Introduction



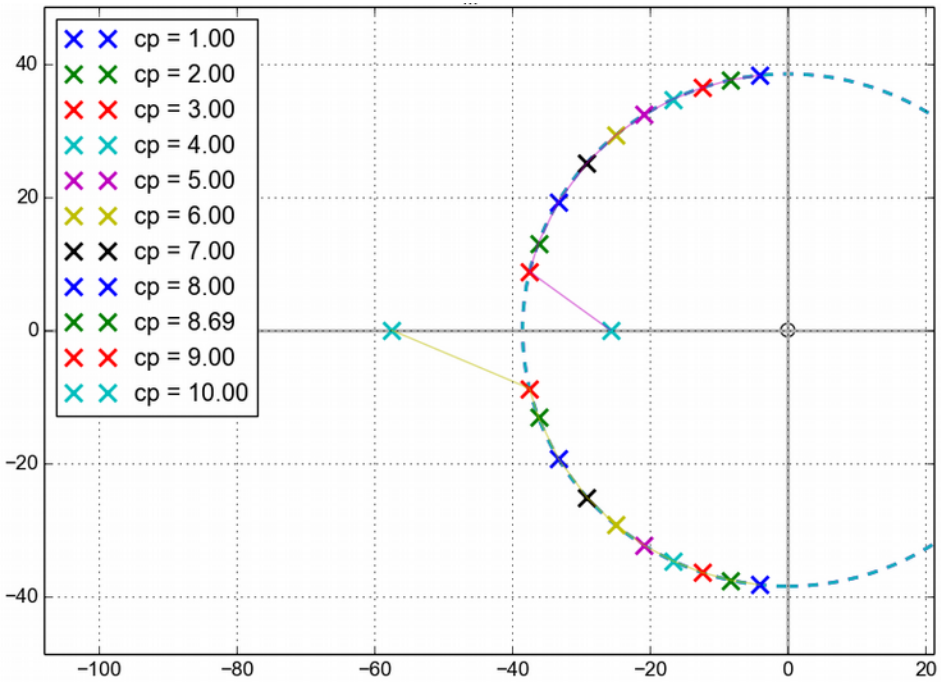
Spring Damper Mass (SDM) model

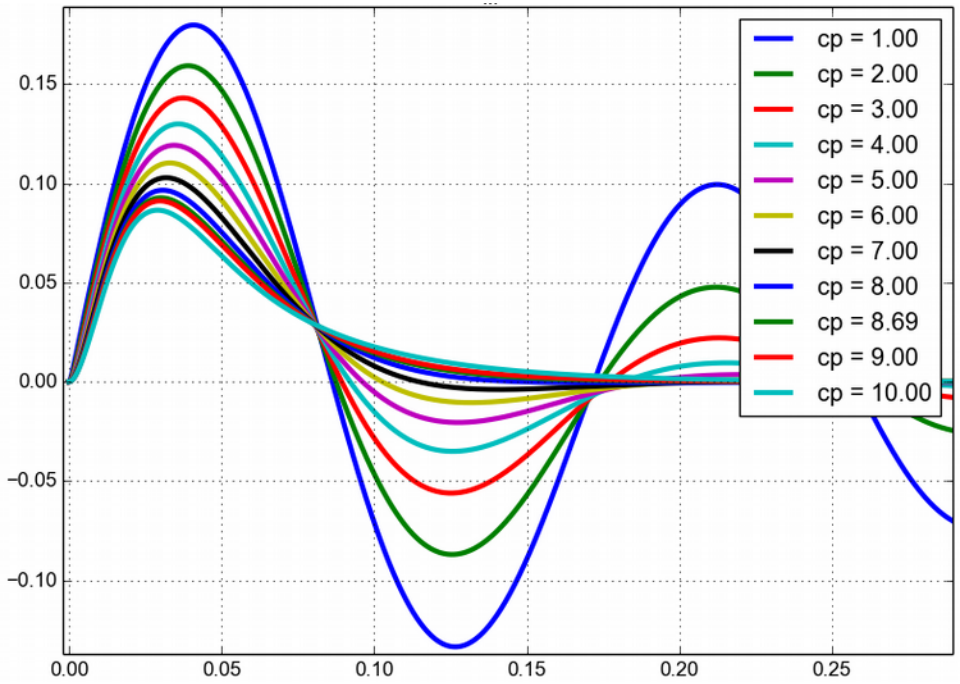


$$y(s) = \frac{1}{\frac{m}{k}s^2 + \frac{c}{k}s + 1}$$

$$\zeta^2 \triangleq \frac{c^2}{4mk} \qquad \omega_0^2 \triangleq \frac{k}{m}$$

SDM model

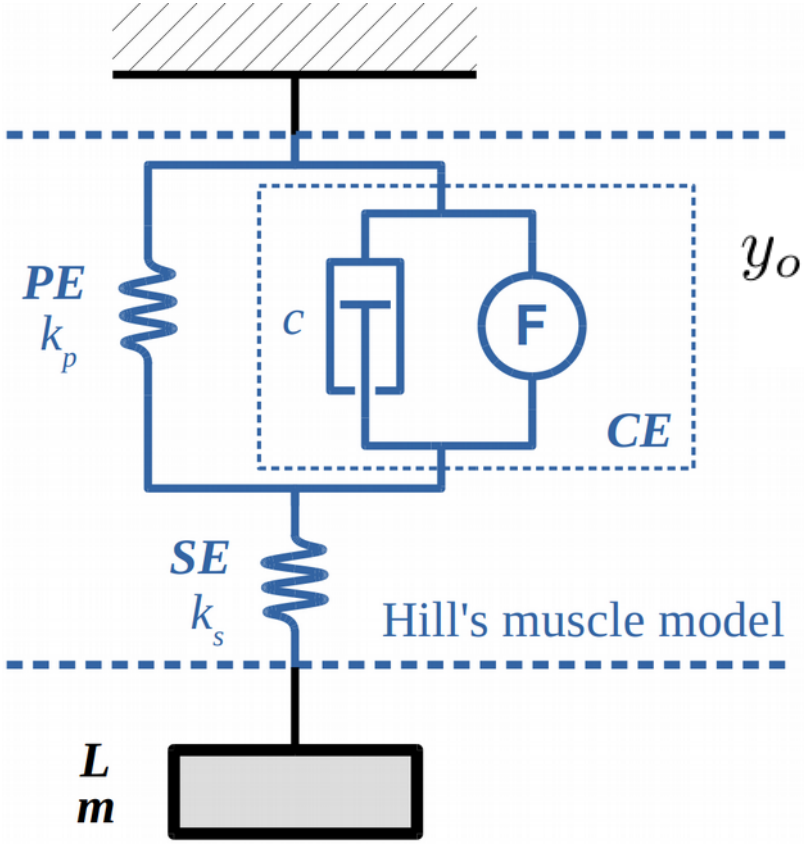

cp = 1.00
cp = 2.00
cp = 3.00
cp = 4.00
cp = 5.00
cp = 6.00
cp = 7.00
cp = 8.00
cp = 8.69
cp = 9.00
cp = 10.00
40
20
0
−20
−40
−100
−80
−60
−40
−20
0
20


cp = 1.00
cp = 2.00
cp = 3.00
cp = 4.00
cp = 5.00
cp = 6.00
cp = 7.00
cp = 8.00
cp = 8.69
cp = 9.00
cp = 10.00
0.15
0.10
0.05
0.00
−0.05
−0.10
0.00
0.05
0.10
0.15
0.20
0.25

SDM model

- The 2nd order SDM does not capture important tendon dynamics.
- It is also not physiologically plausible.

Hill type muscle model

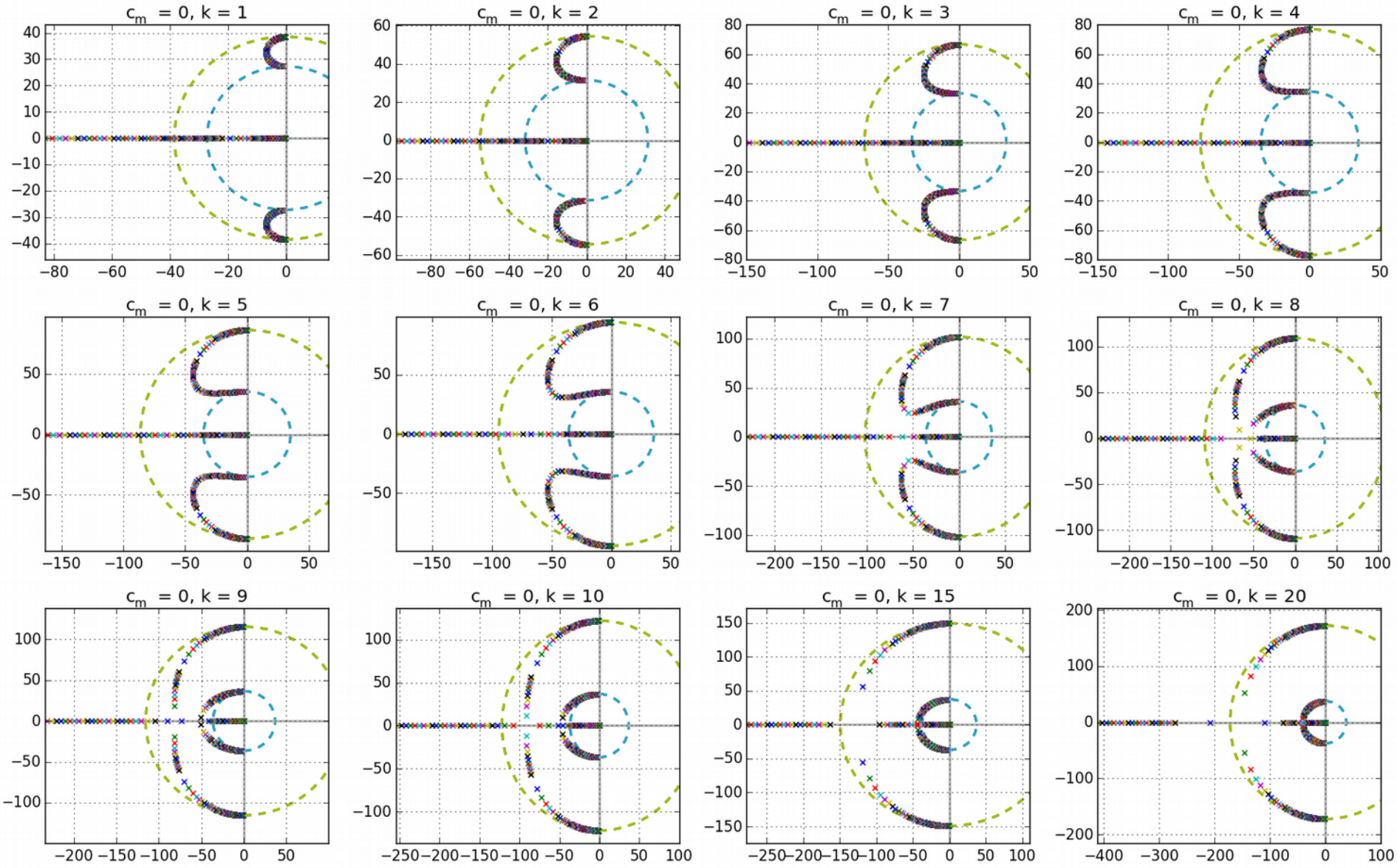


$$y_o(s) = \frac{1}{\frac{cm}{k_s}s^3 + m\frac{k_s+k_p}{k_s}s^2 + cs + k_p}$$

$$\omega_0^2 = \frac{k_p k_s}{m(k_p + k_s)}$$

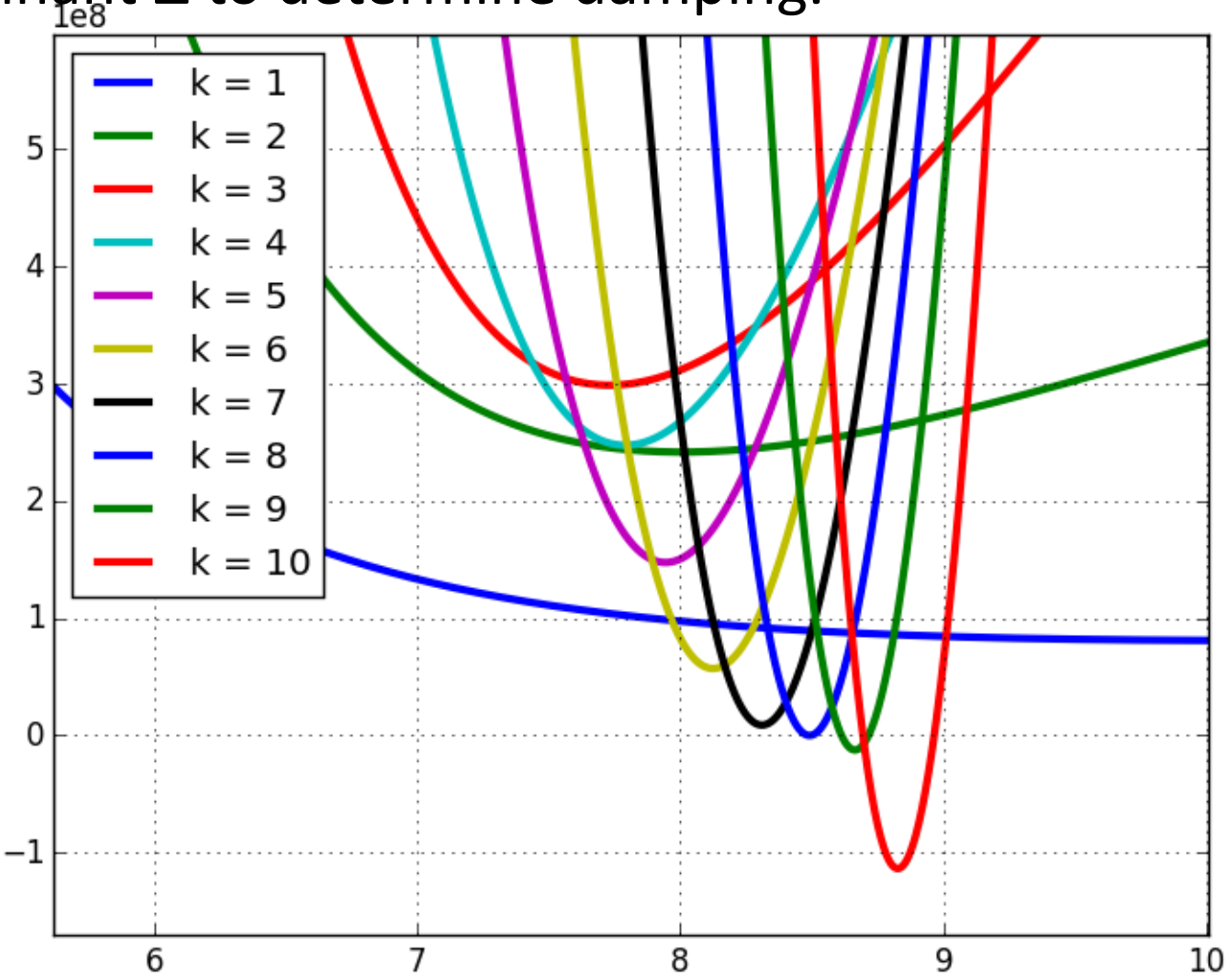
$$k = {}^{k_s}/_{k_p}$$

Hill type muscle model

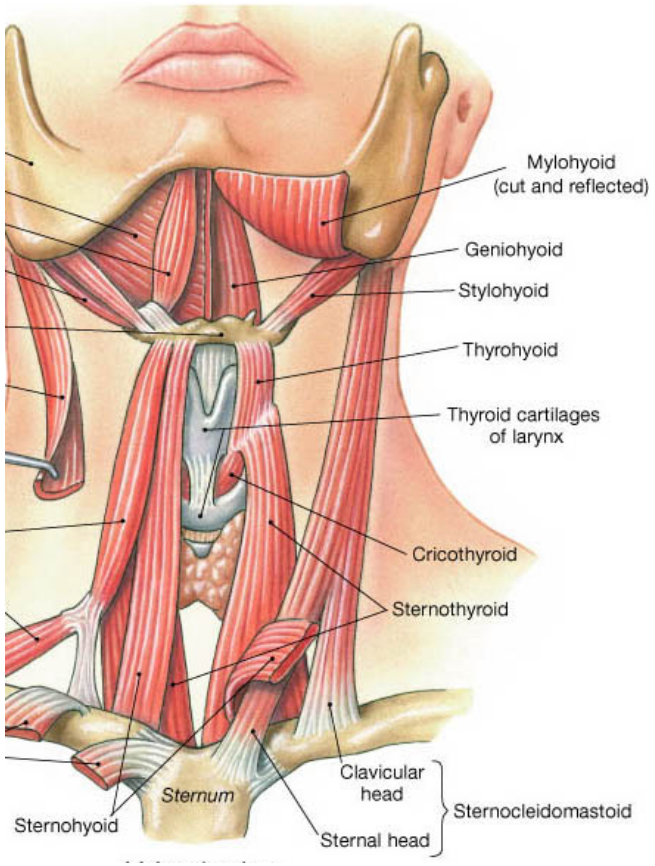


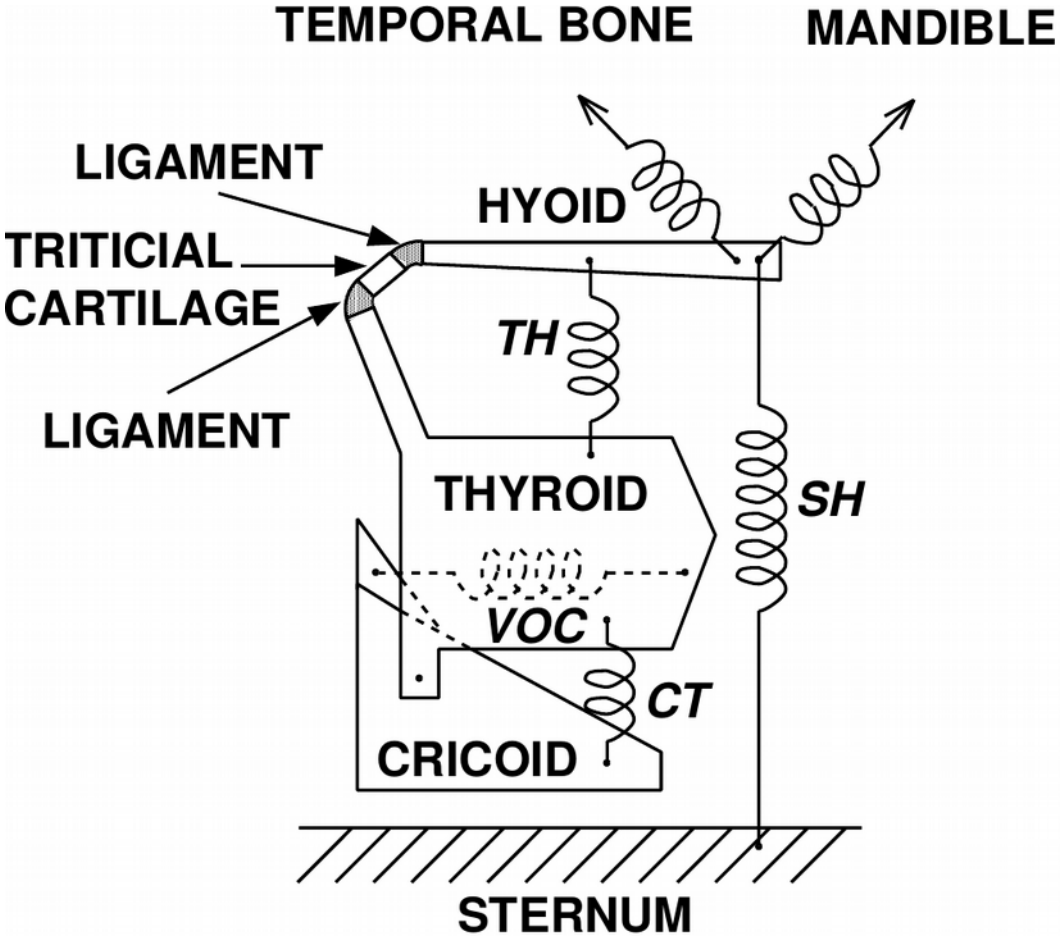
Hill type muscle model

Piovesan *et al.* 2013 used Cardano's formula and the discriminant Δ to determine damping.



Agonist-antagonist pitch production (A2P2) model



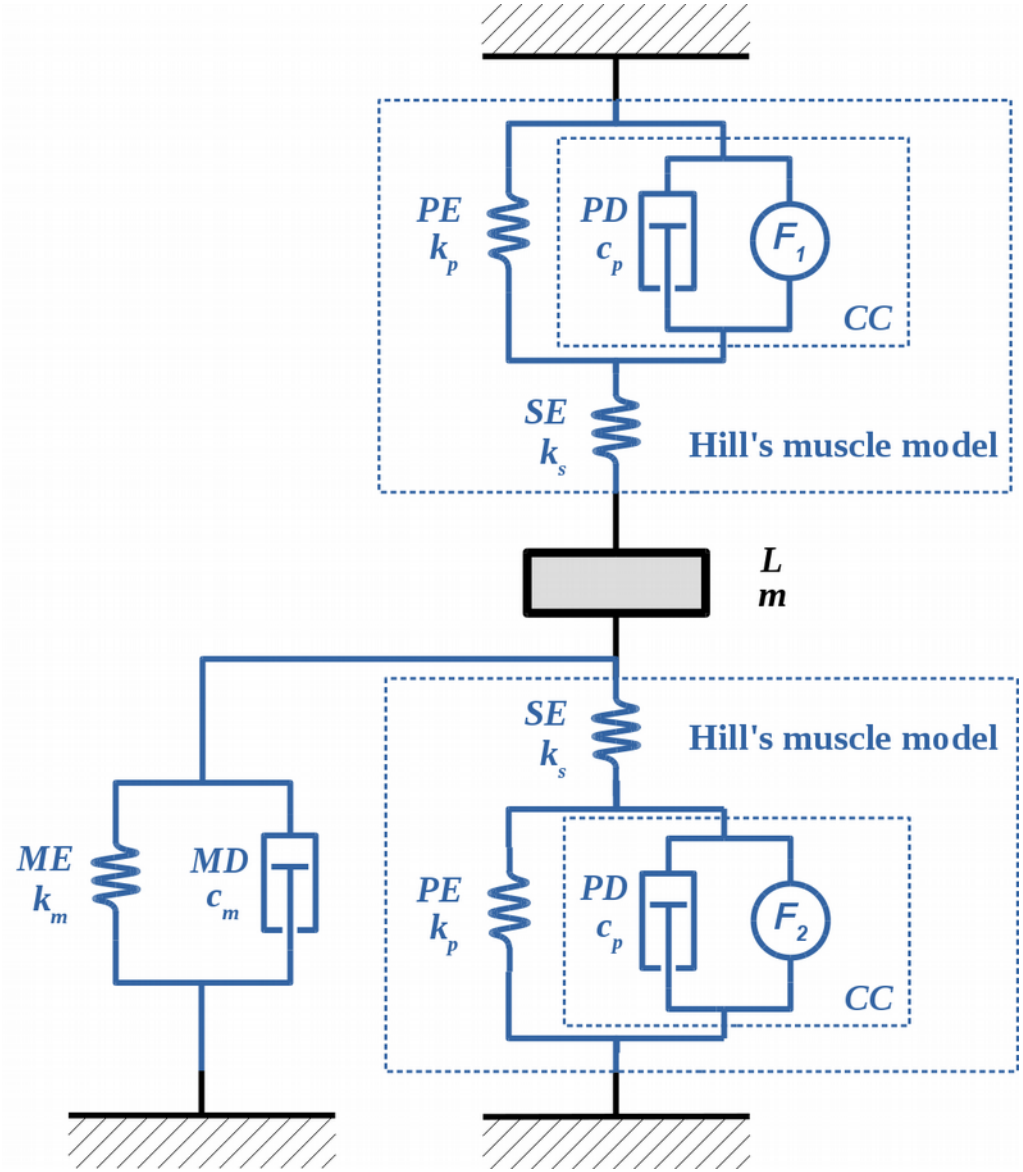


Taken from Fujisaki, H.: The roles of physiology, physics and mathematics in modeling prosodic features of speech. In: Proc. of Speech Prosody. (2006)

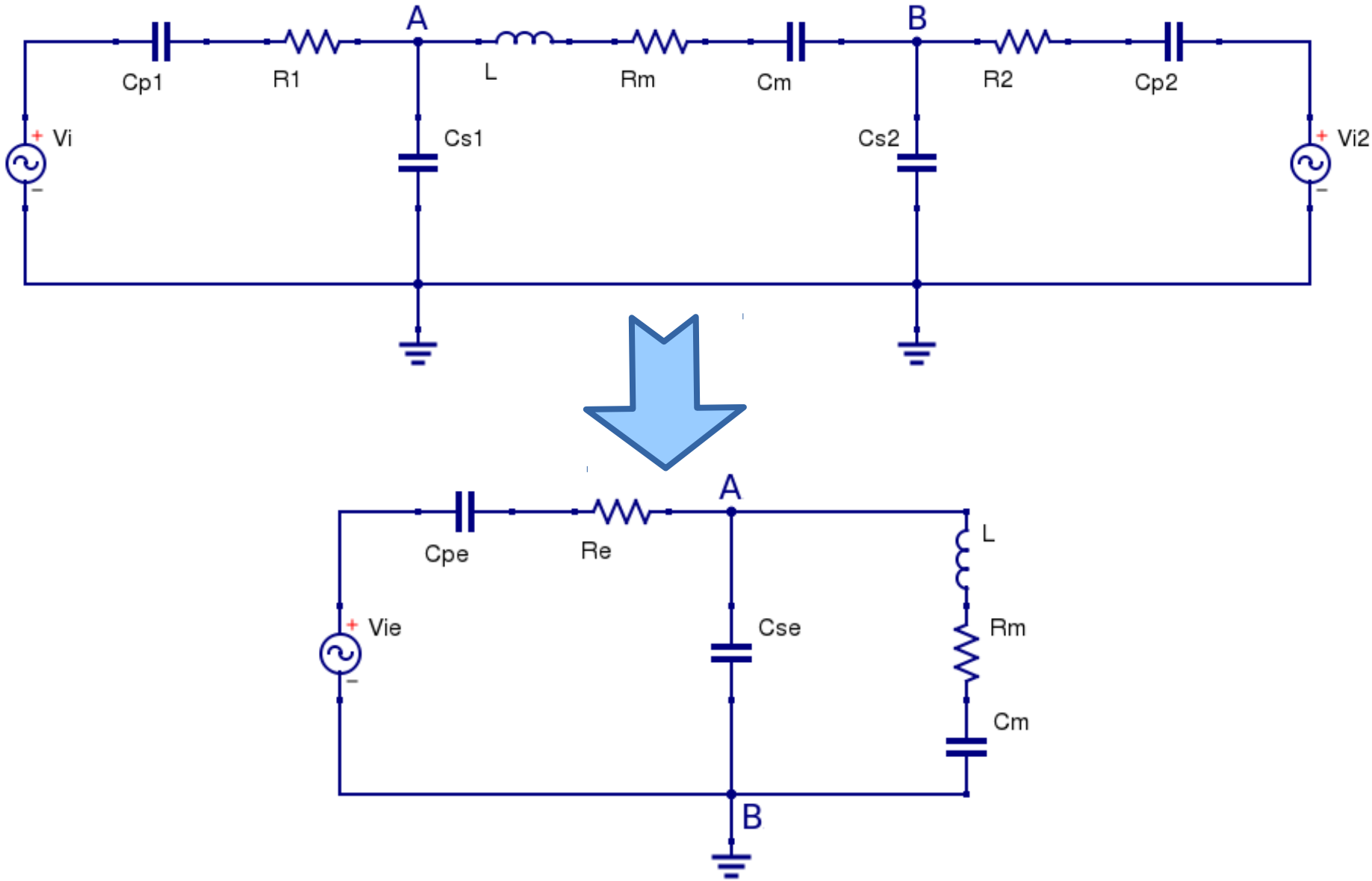
Hill model

- Hill model is the simplest model to capture tendon dynamics.
- It is however underdamped when using physiologically plausible parameters.

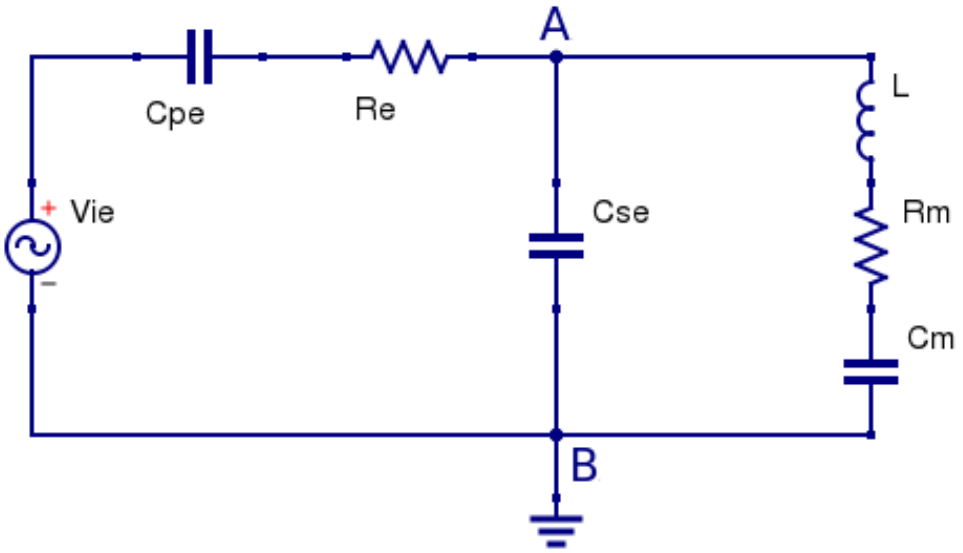
A2P2 model


PE
k_p
PD
c_p
F_1
CC
SE
k_s
Hill's muscle model
L
m
SE
k_s
Hill's muscle model
ME
k_m
MD
c_m
PE
k_p
PD
c_p
F_2
CC

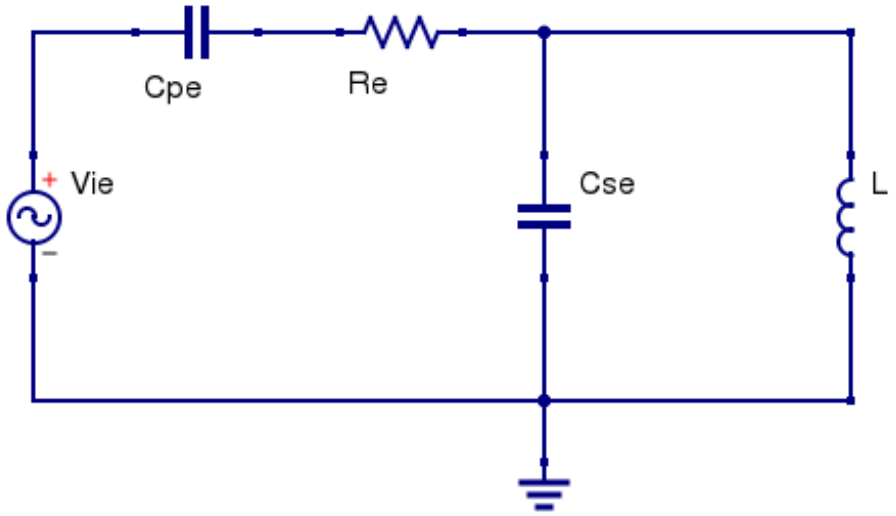
A2P2 model



A2P2 model



Hill model



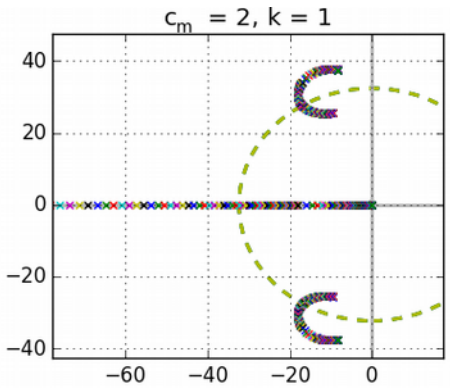
A2P2 model

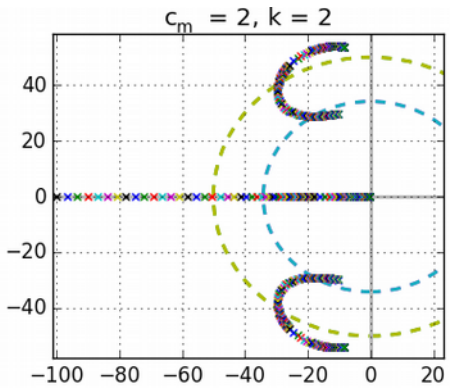
$$y_o(s) = \frac{1}{\frac{c_p m}{k_s} s^3 + \frac{(c_m c_p + m(k_p + k_s)}{k_s} s^2 + \frac{c_m(k_p + k_s) + c_p(k_m + k_s)}{k_s} s + \frac{k_m k_p}{k_s} + k_m + k_p}$$

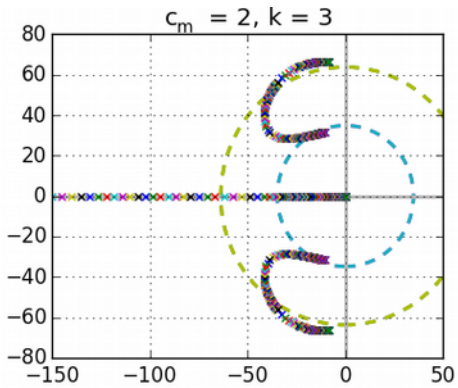
$$\omega_0 = \sqrt{\frac{k_p k_s}{m(k_p + k_s)}}$$

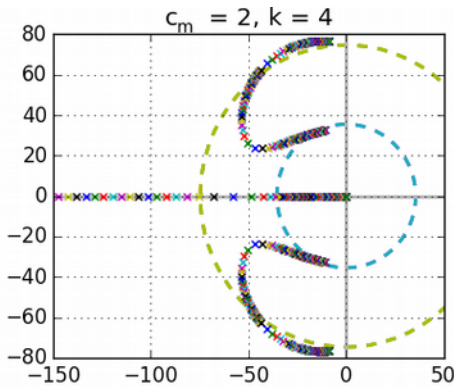
$$\omega_1 = \sqrt{\frac{k_p k_s + k_s^2}{m(k_p + k_s)}}$$

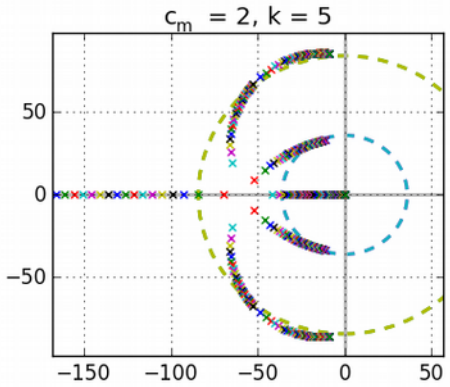
A2P2 model

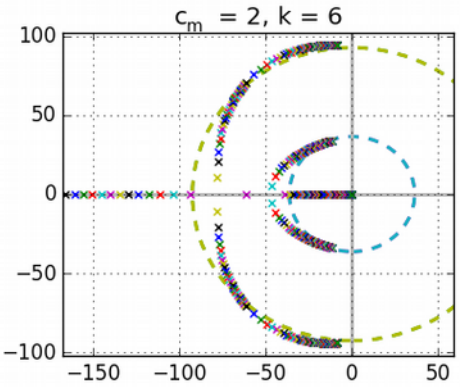

c_m = 2, k = 1

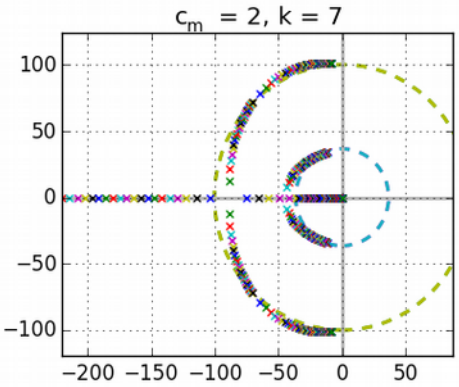

c_m = 2, k = 2

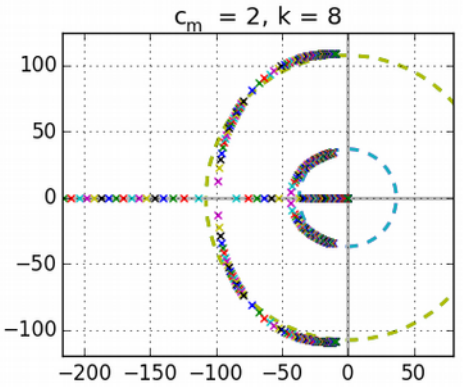

c_m = 2, k = 3

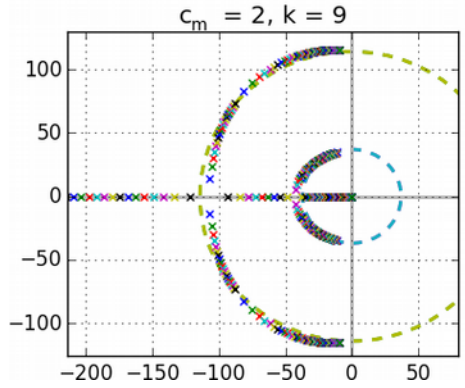

c_m = 2, k = 4

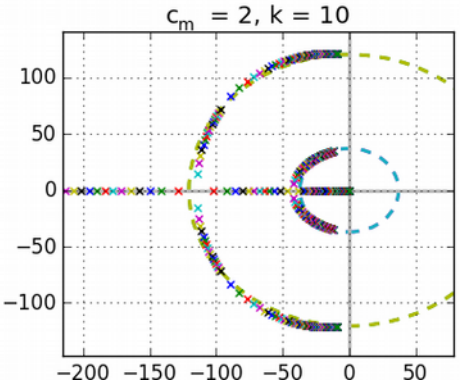

c_m = 2, k = 5

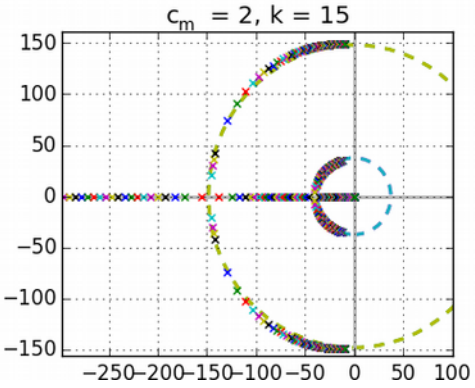

c_m = 2, k = 6

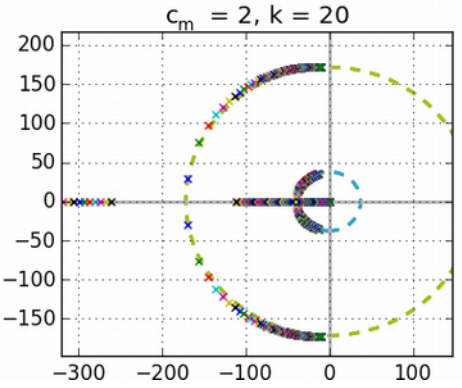

c_m = 2, k = 7


c_m = 2, k = 8

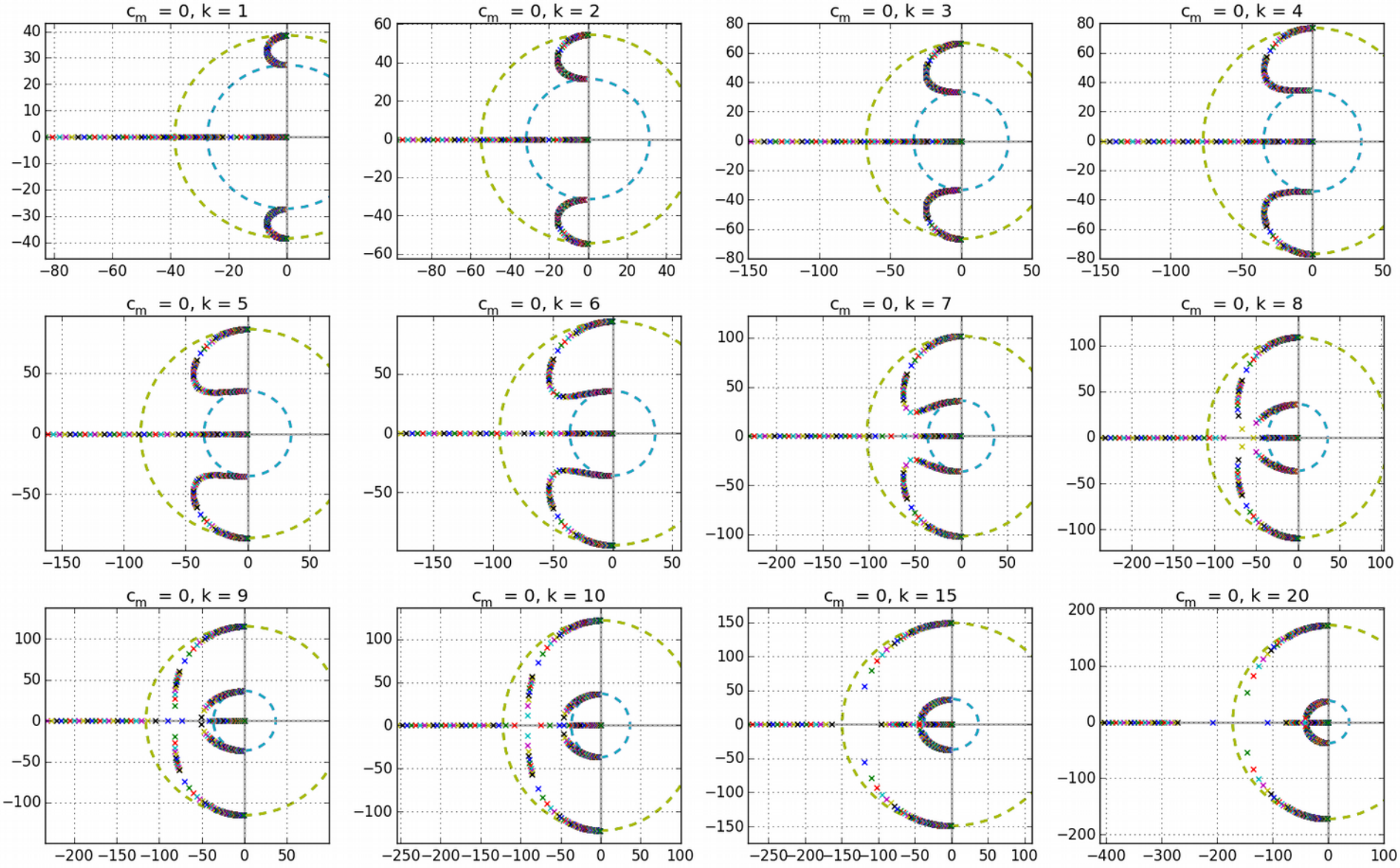

c_m = 2, k = 9


c_m = 2, k = 10

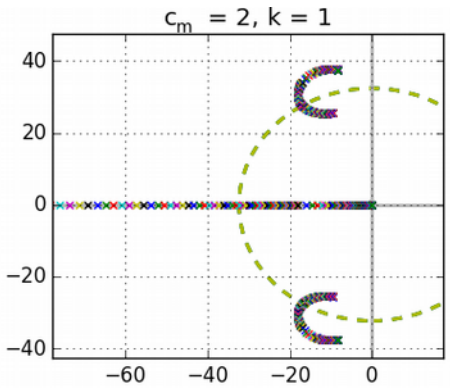

c_m = 2, k = 15

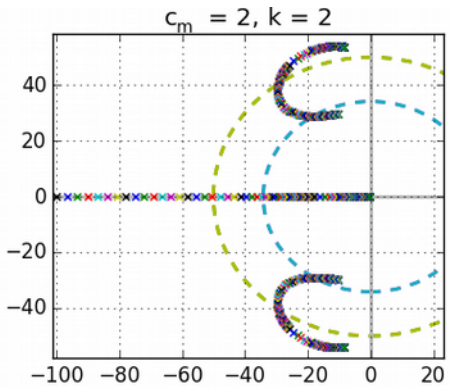

c_m = 2, k = 20

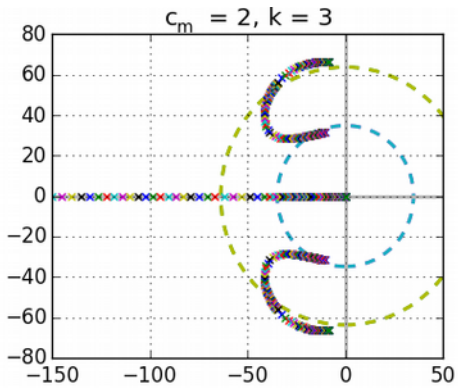
Hill type muscle model

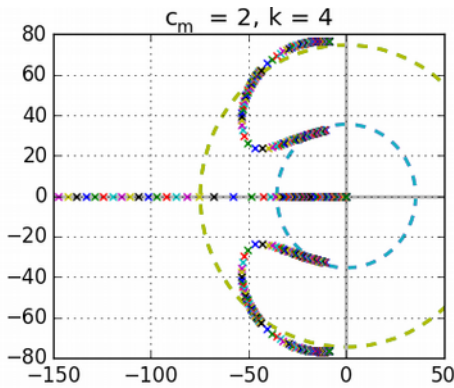


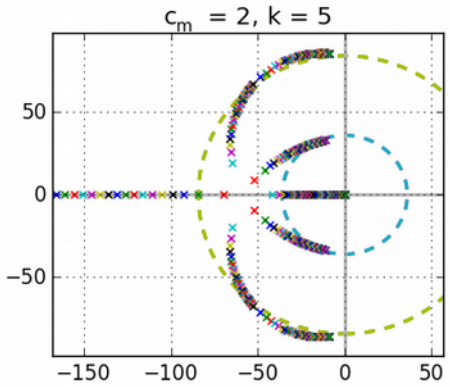
A2P2 model

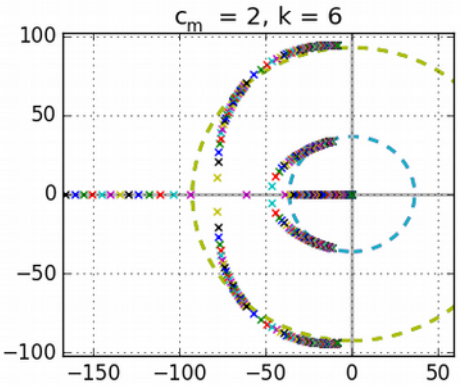

c_m = 2, k = 1

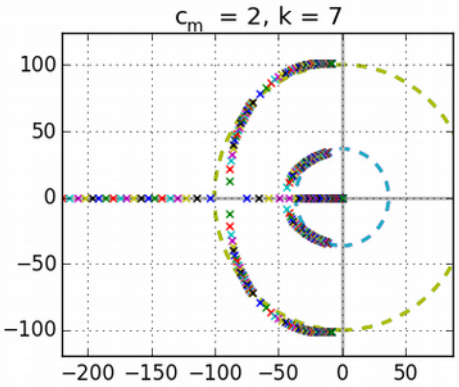

c_m = 2, k = 2

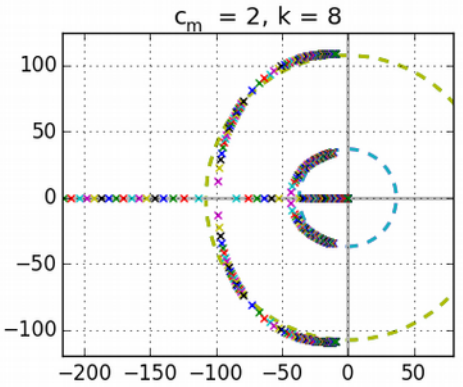

c_m = 2, k = 3

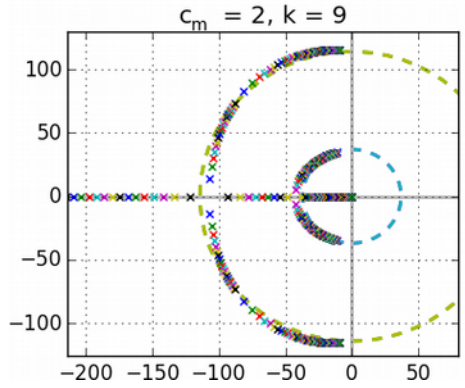

c_m = 2, k = 4

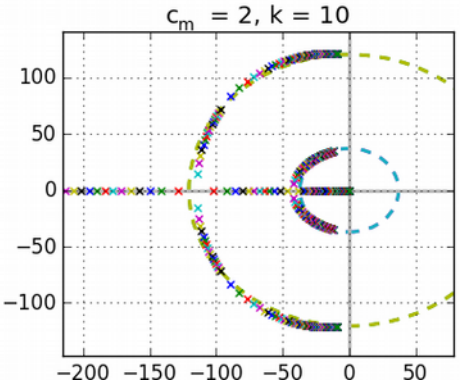

c_m = 2, k = 5

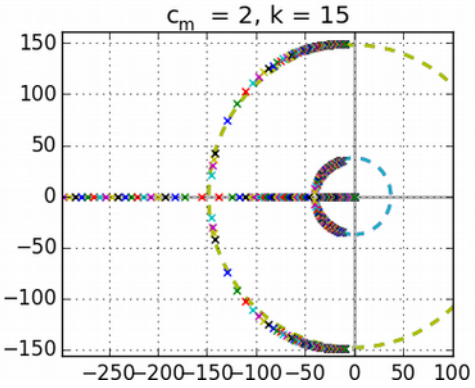

c_m = 2, k = 6

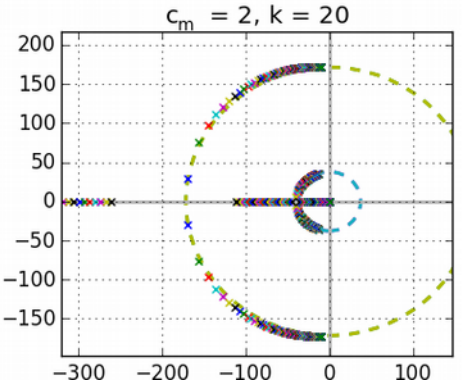

c_m = 2, k = 7


c_m = 2, k = 8

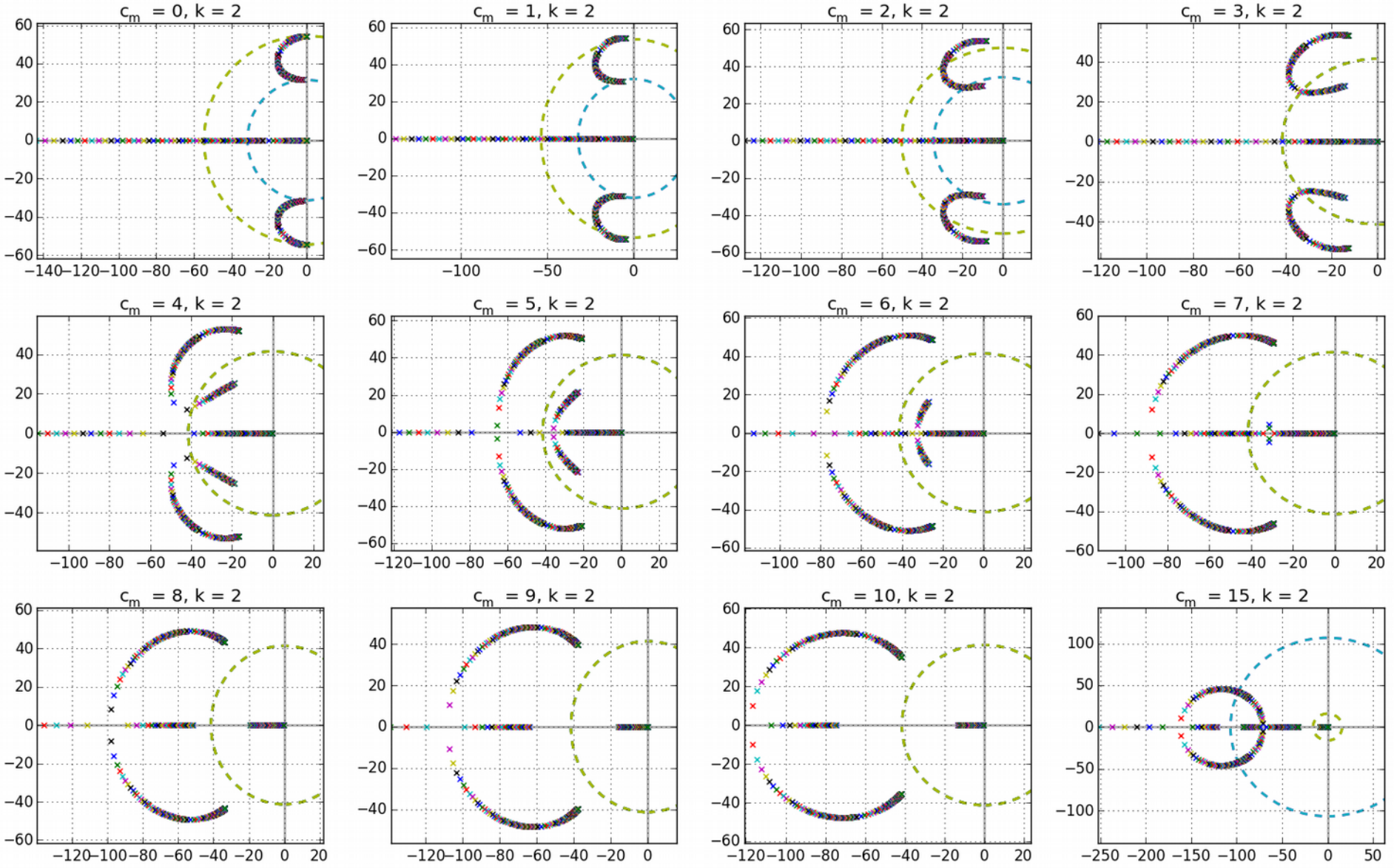

c_m = 2, k = 9


c_m = 2, k = 10


c_m = 2, k = 15


c_m = 2, k = 20

A2P2 model


c_m = 0, k = 2
c_m = 1, k = 2
c_m = 2, k = 2
c_m = 3, k = 2
c_m = 4, k = 2
c_m = 5, k = 2
c_m = 6, k = 2
c_m = 7, k = 2
c_m = 8, k = 2
c_m = 9, k = 2
c_m = 10, k = 2
c_m = 15, k = 2

Conclusion

- We propose an agonist-antagonist pitch production (A2P2) model to capture the opposing muscle physiological environment of pitch production.
- A simplified version of the model exhibits critical damping already for $k = 5$, for a thyro-cricoid joint damping of $c_m = 2$.
- For higher thyro-cricoid damping the model's damping range is even further increased.
- The model grants physiological plausibility to the use of critically damped, higher order system models in intonation modelling.

Acknowledgement

- The work was funded by the Swiss National Science Foundation (SNSF) under the SCOPES program:

  “SP2 - SCOPES Project on Speech Prosody”

An agonist-antagonist pitch production (A2P2) model

Branislav Gerazov and Philip N. Garner

Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University of Skopje, Macedonia

Idiap Research Institute, Martigny, Switzerland