

# An Investigation of Muscle Models for Physiologically Based Intonation Modelling

Branislav Gerazov, *Graduate Student Member, IEEE*, and Philip N. Garner, *Senior Member, IEEE*

**Abstract**—Prosody is a crucial aspect of the speech signal and its modelling is of great importance for various speech technologies. Intonation models based on physiology rely on an accurate model of muscle activation. Although most of them are based on the spring-damper-mass (SDM) muscle model, the more complex Hill type model offers a more accurate representation of muscle dynamics. In this paper we analyse and compare these two muscle models and discuss the benefits and disadvantages they bring. This research is a part of an on-going effort to develop an improved intonation model.

**Key words**—Intonation modelling, muscle models, physiology, biomechanics

## I. INTRODUCTION

Prosody is a multidimensional phenomenon comprising the intonation, energy, and duration contours of the speech signal. It is the carrier of linguistic information, including sentence structure, mode of enunciation, focus and contrast, lexical stress, and even meaning. Prosody also serves to carry paralinguistic information, such as gender, age, personality, ethnic background, emotions, and physiological state.

Although prosody carries a wealth of information, it has been only minimally integrated in speech technology systems. In Text to Speech synthesis (TTS), prosody is crucial for generating intelligible speech output. However these systems use an averaged, worn-down version of prosody that sounds bland and not quite natural. On the other hand, prosody is completely ignored in Automatic Speech Recognition (ASR) systems, except in tonal languages, where it serves to carry linguistic meaning. This suboptimal treatment of prosody is due to our lack of understanding of how prosody really works, which is manifest by the lack of a satisfactory modelling framework. This problem has been emphasized recently with the shift of scientific focus on the areas of Speech Emotion Recognition (SER), emotional speech synthesis, and emphatic human-machine dialogue, all of which rely heavily on prosody.

The importance of prosody in TTS systems has been the driving force in the development of prosody models [1], most of which focus on intonation. Intonation models can be roughly split into two groups: 1) surface models, which model the pitch contour directly, and 2) physiological models, that model the underlying mechanisms of pitch production by

This work has been conducted with the support of the Swiss National Science Foundation SCOPES programme, in the framework of the SP2: SCOPES Project on Speech Prosody.

B. Gerazov is with the Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University – Skopje (FEEIT), Skopje, Macedonia

P. N. Garner is with the Idiap Research Institute, CH-1920 Martigny, Switzerland.

directly or indirectly incorporating physiological constraints. While there is a plethora of surface intonation models, physiologically based models are not as numerous. The most famous model from this group is the Fujisaki Command-Response (CR) model [2]. The CR model treats the pitch contour as a sum of global, phrase components, and local, accent components. Both components are output from a 2<sup>nd</sup> order critically damped system, which is excited either by impulses or by step functions, respectively. The 2<sup>nd</sup> order system models the dependence of pitch on laryngeal muscle activation [3]. Other authors have shown that using higher order system models is beneficial to intonation modelling, e.g. the quantitative target approximation qTA model [4], which uses pitch targets as a physiological force function input to a 3<sup>rd</sup> order system to generate the surface pitch contours.

Recently, we have proposed an intonation model that offers improvements in consistency and physiological plausibility, while maintaining comparable performance across speakers and languages [5], [6]. Our model is based on the decomposition of the pitch contour into gamma shaped elementary atoms, which can be used to approximate the impulse response of a  $n$ -th order critically damped system. Similarly to Prompon *et al.* [4], our results have also shown improved model performance when using higher, up to 6<sup>th</sup> order impulse responses as atoms. These findings have prompted us to analyse in more detail the various muscle activation models. There are different muscle models suggested in literature, which go from very detailed ones - modelling the internal mechanics of the muscle fibre, to more general ones - modelling only the output to a given input of the muscle as a whole [7]. In this paper we present an analysis that is part of an on-going research effort, in which we take a closer look at the two most famous muscle models: the spring-damper-mass (SDM) model and the Hill type model.

## II. THE SPRING-DAMPER-MASS (SDM) MUSCLE MODEL

It is a necessary practice to simplify the complex non-linear mechanics of muscle for the purpose of their analysis. The spring-damper-mass (SDM) model shown in Fig. 1 is the simplest model of muscle activation. It's critically damped 2<sup>nd</sup> order response served as a basis for Fujisaki *et al.* [2] to develop the CR intonation model.

The output movement  $y(t)$  of the mass is given in Laplace domain in (1), as a function of the input movement  $x(t)$ , which is caused by an input force  $f_i(t)$  [8].

$$Y(s) = \frac{sy(0) + y'(0) + 2\zeta\omega_0y(0) + \omega_0^2X(s)}{s^2 + 2\zeta\omega_0s + \omega_0^2} \quad (1)$$

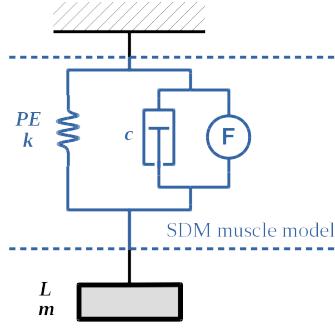


Fig. 1. A 2<sup>nd</sup> order spring-damper-mass (SDM) muscle model.

Here  $\zeta$  is the damping ratio, and  $\omega_0$  is the undamped natural frequency of the mechanical system [9]. They are defined as:

$$\zeta^2 \triangleq \frac{c^2}{4mk} , \quad \omega_0^2 \triangleq \frac{k}{m}. \quad (2)$$

If we assume critical damping ( $\zeta = 1$ ), steady state initial conditions ( $y(0) = 0$  and  $y'(0) = 0$ ), and an impulsive driving force ( $X(s) = 1$ ), from (1) we have:

$$Y(s) = \frac{\omega_0^2}{s^2 + 2\omega_0 s + \omega_0^2} = \frac{\omega_0^2}{(s + \omega_0)^2} \quad (3)$$

the solution of which is:

$$y(t) = \omega_0^2 t e^{-\omega_0 t} u(t) \quad (4)$$

where  $u(t)$  is the Heaviside step function. This impulse response function is used in the CR model.

The zero-pole diagram and the corresponding impulse responses for a sweep of values of  $c \in [1, 10]$ , for  $k = 178$  and  $m = 0.12$  are given in Fig. 2. The parameters were taken from Piovesan *et al.* [10] and represent the elasticity of the elbow muscles and inertia of the underarm. As we can see, although not that evident in the plots of the impulse response, the system reaches critical damping only for  $c = 10$ , which is at the extreme end of the physiologically plausible range.

### III. THE HILL TYPE MUSCLE MODEL

The three-element Hill muscle model is shown in its Poynting-Thomson (PT) form [7] in Fig. 3, it comprises:

- a *contractile element* (CE) - representing the active force generated in the muscle tissue through its contraction. This component can be modelled as a parallel circuit comprising a force generator and a damper,
- a *series elasticity* (SE) - representing the elasticity of the tendons that connect the muscle to the bone,
- a *parallel elasticity* (PE) - representing the elasticity of the muscle tissue itself.

To derive the characteristic function of this model we will use the impedance electro-mechanical analogy [11]. This analogy allows us to use methods and tools developed for the analysis of electrical circuits to analyse the dynamics of a mechanical system. The equivalent electrical circuit is shown in Fig. 4. Here, the resistor  $R = c$  represents the damping of the CE, the inductance  $L = m$  represents the mass load,

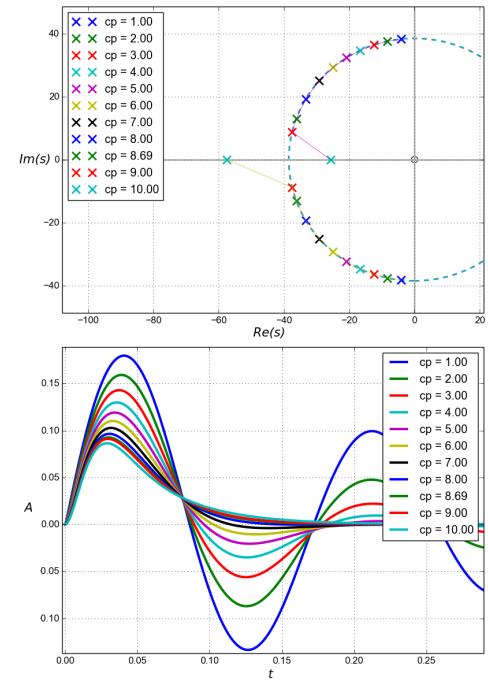


Fig. 2. Zero pole diagram and impulse response of the 2<sup>nd</sup> order spring-damper-mass (SDM) muscle model for a sweep of values of  $c \in [1, 10]$ , for  $k = 178$  and  $m = 0.12$ .

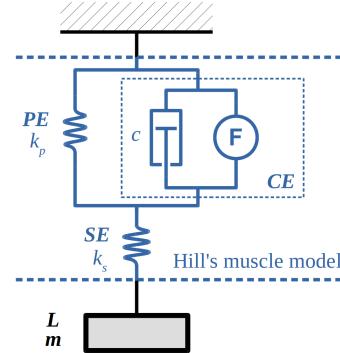


Fig. 3. Hill's three element model.

and the two capacitors  $C_p = 1/k_p$  and  $C_s = 1/k_s$  represent the elasticity of the PE and SE, respectively.

We are ultimately interested in the movement that the muscle generates we want to track the displacement of the load, which is analogous to the charge in the inductance. If we solve for the charge in the inductor assuming steady state

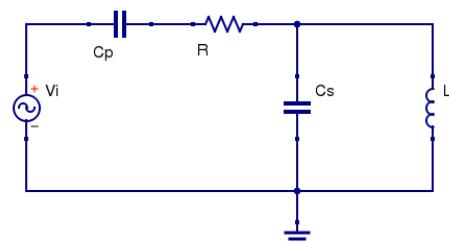


Fig. 4. Equivalent electrical circuit to Hill's three element model with mass load.

initial conditions, in the Laplace domain [9] we obtain (5), which is equivalent to (6) in the mechanical domain for  $f_i(t) = \delta(t)$ .

$$q_0(s) = \frac{i_o(s)}{s} = \frac{1}{RLC_s s^3 + L \frac{C_s + C_p}{C_p} s^2 + Rs + \frac{1}{C_p}} v_i(s) \quad (5)$$

$$y_o(s) = \frac{1}{\frac{cm}{k_s} s^3 + m \frac{k_s + k_p}{k_s} s^2 + cs + k_p} \quad (6)$$

From the circuit shown in Fig. 4 we can find the resonant frequency of the mechanical system. This can be done by finding the input impedance  $Z_i(s)$ , substituting  $s = j\omega$ , finding the imaginary part and equalling it to zero [9].

$$Z_i(s) = \frac{v_i(s)}{i_i(s)} = R + \frac{1}{sC_p} + \frac{sL}{1 + s^2 LC_s} \quad (7)$$

$$X(\omega_0) = \frac{\omega_0 L}{1 - \omega_0^2 LC_s} - \frac{1}{\omega_0 C_p} = 0 \quad (8)$$

$$\omega_0^2 = \frac{1}{L(C_s + C_p)} \quad (9)$$

or in the mechanical domain:

$$\omega_0^2 = \frac{k_p k_s}{m(k_p + k_s)} \quad (10)$$

We can use (5) and (6) to find the characteristic equations of the system in the two domains. By assuming zero-input and steady-state conditions we have:

$$RLC_s s^3 + L \frac{C_s + C_p}{C_p} s^2 + Rs + \frac{1}{C_p} = 0 \quad (11)$$

$$\frac{cm}{k_s} s^3 + m \frac{k_s + k_p}{k_s} s^2 + cs + k_p = 0 \quad (12)$$

Finally, we can rewrite (12) by substituting in the resonant frequency  $\omega_0$  from (10) and introducing the ratio of the elasticity of the SE and PE  $k = k_s/k_p$  as:

$$\frac{c}{k(1+k)} s^3 + s^2 + \frac{c}{k_p} \omega_0^2 s + \omega_0^2 = 0. \quad (13)$$

To derive the damping ratio  $\zeta$  of this system we can parametrize the 3<sup>rd</sup> order polynomial into:

$$(s + \alpha \omega_0)(s^2 + 2\zeta \omega_0 s + \omega_0^2) = 0. \quad (14)$$

However, this leads to an overdetermined system of nonlinear equations that has no solutions, except the trivial one. This points to the necessity of taking a different approach in the analysis.

Piovesan *et al.* [10] did a thorough analysis of (12) used Cardano's formula and the discriminant  $\Delta$  to clarify the connection between the system damping and the elasticity ratio  $k$ . A  $\Delta > 0$  means that the system is underdamped – it has two complex conjugate poles, and oscillates; for  $\Delta = 0$  the system is critically damped – the complex poles become real and equal; for  $\Delta < 0$  the system is overdamped – all of the

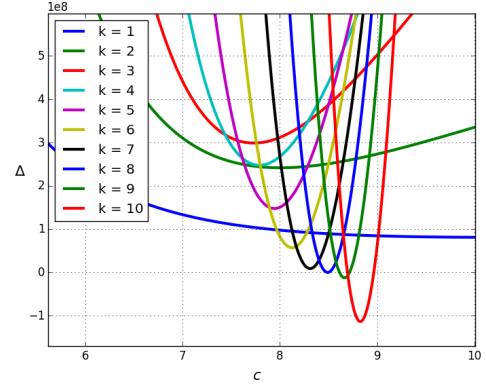


Fig. 5. Value of the discriminant  $\Delta$  in respect to the CE damping  $c$  for different values of  $k = k_s/k_p$ .

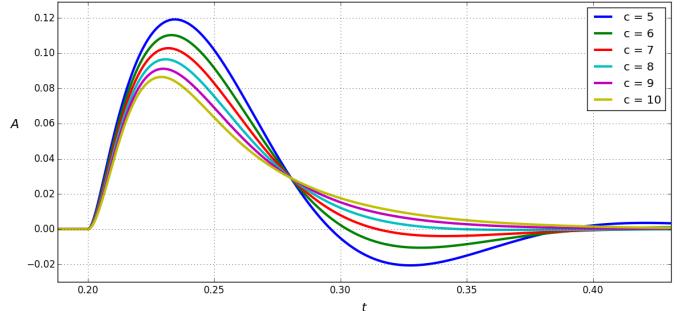


Fig. 6. Numerical simulation of 3<sup>rd</sup> order impulse response using elbow parameters [10], for a sweep of values of the muscle damping  $c$ .

poles are real, but different. In our case, if we define  $d = c/k_p$ , for the discriminant  $\Delta$  we obtain:

$$q = \frac{(3k+3)d^2 \omega_0^2 - k^2 - 2k - 1}{9d^2} \quad (15)$$

$$r = \frac{(9k^2 - 9k - 18)d^2 \omega_0^2 - 2k^3 - 6k^2 - 6k - 2}{54d^3} \quad (16)$$

$$\Delta = q^3 + r^2 \quad (17)$$

A numerical analysis of (17) using the physiological parameters for the elbow joint [10], yields the plots shown in Fig. 5. We can see that indeed the damping depends on the ratio  $k$ , and that the system reaches critical damping only for  $k \geq 8$ . The impulse response of the system for  $k = 10$  and  $c \in [5, 10]$  is shown in Fig. 6. These plots also show that the response reaches critical damping for  $8 < c < 9$ .

#### IV. COMPARISON AND DISCUSSION

The two presented muscle models are of different complexity and thus capture the physiology of muscle activation to a different extent. It is necessary to analyse the tradeoff that each of these models offers, when choosing which one to use for the application of intonation modelling. Our discussion will be guided according to the following questions:

- 1) Does the SDM model oversimplify muscle dynamics?
- 2) Is using a critically damped Hill model physiologically plausible?
- 3) Do we need a more complex model?

### A. SDM oversimplifying muscle dynamics

Researchers mostly use 2<sup>nd</sup> order systems to model muscle activation, in favour to 3<sup>rd</sup> order ones, under the assumption that the tendon's stiffness is much larger than that of the stiffness of the muscle tissue itself. Even though SE stiffness is a multiple of PE stiffness, this claim does not hold entirely, because it is never beyond an order of magnitude greater. Indeed, research has shown that  $k < 8$  for the calf muscle [12],  $k < 10$  for the elbow muscles at full activation [13], and it can be as little as  $k = 1$  for muscles in the palm at full activation [14].

Kistemaker *et al.* have used a 2<sup>nd</sup> order SDM model to replicate the response of a 3<sup>rd</sup> order Hill model [13]. They have found that although reasonable accuracy can be obtained for the initial (rising) portion of the impulse response, the SDM model fails to adequately capture the whole length of the response. Others, have found a maximum correlation of 0.85 between observed real muscle responses and simulated ones with the SDM model [15].

Moreover, the major flaw of the SDM model is the complex non-linear relationship between its parameters and those of the Hill model. This means that not only is the estimated stiffness and damping very different from the actual modelled parameters, and outside of the physiologically plausible range, but also that they are very sensitive to measurement errors [13].

### B. Critically damped Hill model

The linearised Hill model neglects the important force-length-velocity non-linear relationship of real muscles, as well as the dynamical properties originating from activation dynamics, short range stiffness and history dependence of muscle force production. Still, it is the simplest model that takes into account the essential interaction arising from the stiffness of the tendon [13], and is a faithful representation of muscle-tendon mechanics [10]. This can be the reason behind the increased performance of the qTA when using a 3<sup>rd</sup> order system response [4].

As was shown, the increased complexity of the Hill model precludes a compact mathematical solution. This can be mitigated by taking the assumption that the system is critically damped. However, Piovesan *et al.* have shown that conditions for critical damping are minuscule when one uses physiologically plausible model parameters [10]. Over most of the parameter range the Hill system is underdamped. This suggests that using a critically damped Hill model is not justified in physiology.

### C. More complex models

On one hand, the underdamped behaviour of the Hill type model can be used to explain why humans use position control strategy, i.e. the central nervous system (CNS) activates opposing muscles to indirectly control force [16]. On the other, measurements of isotonic contractions have not shown oscillations in single muscle fibre contraction, which might suggest that the Hill model is not enough.

Indeed, considering the complex interaction between several muscles that drives the changes in  $F_0$ , we can argue that a

more complex mechanical model is needed. Such a model should also capture the mechanics of the joints and the vocal folds themselves. It is to be expected that such a model would also have increased order, which might explain the increase in performance that we have seen when using 6<sup>th</sup> order atoms in our intonation model [6].

## V. CONCLUSION

We can conclude that although the SDM model has been successfully used to model intonation, it is too simple to capture the basic mechanics of muscle activation. In that sense, better performance can be expected when using higher order models. The Hill type model is the next step in complexity, offering improved modelling of muscle-tendon dynamics. However, the model exhibits underdamped behavior which cannot be parametrized easily. Finally, the complexity of the physiology of pitch production hints towards a necessity of even more complex models, which necessitates further investigation.

## REFERENCES

- [1] Jan van Santen, Tanya Mishra, and Esther Klabbers, "Prosodic processing," in *Springer Handbook of Speech Processing*, pp. 471–488. Springer, 2008.
- [2] Hiroya Fujisaki and Shigeo Nagashima, "A model for the synthesis of pitch contours of connected speech," Tech. Rep., Engineering Research Institute, University of Tokyo, 1969.
- [3] Hiroya Fujisaki, "The roles of physiology, physics and mathematics in modeling prosodic features of speech," in *Proc. of Speech Prosody*, 2006.
- [4] Santitham Prom-on, Yi Xu, and Bundit Thipakorn, "Modeling tone and intonation in mandarin and english as a process of target approximation," vol. 125, pp. 405–424, January 2009.
- [5] Pierre-Edouard Honnet, Branislav Gerazov, and Philip N. Garner, "Atom decomposition-based intonation modelling," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, April 2015.
- [6] Branislav Gerazov, Pierre-Edouard Honnet, Aleksandar Gjoreski, and Philip N. Garner, "Weighted correlation based atom decomposition intonation modelling," in *Proceedings of Interspeech*, Sep 2015.
- [7] Vladimir Zatsiorsky and Boris Prilutsky, *Biomechanics of skeletal muscles*, Human Kinetics, 2012.
- [8] Philip N. Garner, "A derivation of a second order damped system," <http://www.idiap.ch/~pgarner/appendices/damping.pdf>.
- [9] Roland E Thomas, Albert J Rosa, and Gregory J Toussaint, *The analysis and design of linear circuits*, 7th Ed., Wiley Publishing, 2012.
- [10] D. Piovesan, A. Pierobon, and F. A. Mussa Ivaldi, "Critical damping conditions for third order muscle models: implications for force control," *J Biomech Eng*, vol. 135, no. 10, pp. 101010, Oct 2013.
- [11] Leo Beranek, *Acoustics sound fields and transducers*, Academic Press, S.l, 2012.
- [12] I. D. Loram, M. Lakie, I. Di Giulio, and C. N. Maganaris, "The consequences of short-range stiffness and fluctuating muscle activity for proprioception of postural joint rotations: the relevance to human standing," *J. Neurophysiol.*, vol. 102, no. 1, pp. 460–474, Jul 2009.
- [13] Dinant A. Kistemaker and Leonard A. Rozendaal, "in vivo; dynamics of the musculoskeletal system cannot be adequately described using a stiffness-damping-inertia model," *PLoS ONE*, vol. 6, no. 5, pp. e19568, 05 2011.
- [14] C. S. Cook and M. J. McDonagh, "Measurement of muscle and tendon stiffness in man," *Eur J Appl Physiol Occup Physiol*, vol. 72, no. 4, pp. 380–382, 1996.
- [15] Toshihiko Tsujii, Barbara Attardi, and Stephen J. Winters, "Regulation of  $\alpha$ -subunit mRNA transcripts by pituitary adenylate cyclase-activating polypeptide (pacap) in pituitary cell cultures and t3-1 cells," *Molecular and Cellular Endocrinology*, vol. 113, no. 2, pp. 123 – 130, 1995.
- [16] Florin Popescu, Joseph M Hidler, and W Zev Rymer, "Elbow impedance during goal-directed movements," *Experimental brain research*, vol. 152, no. 1, pp. 17–28, 2003.