An agonist-antagonist pitch production model

Branislav Gerazov[1] and Philip N. Garner[2]

[1] Faculty of Electrical Engineering and Information Technologies, University of Ss. Cyril and Methodius in Skopje, Macedonia
gerazov@feit.ukim.edu.mk
[2] Idiap Research Institute, Martigny, Switzerland
phil.garner@idiap.ch

Abstract. Prosody is a phenomenon that is crucial for numerous fields of speech research, accenting the importance of having a robust prosody model. A class of intonation models based on the physiology of pitch production are especially attractive for their inherent multilingual support. These models rely on an accurate model of muscle activation. Traditionally they have used the 2nd order spring-damper-mass (SDM) muscle model. However, recent research has shown that the SDM model is not sufficient for adequate modelling of the muscle dynamics. The 3rd order Hill type model offers a more accurate representation of muscle dynamics, but it has been shown to be underdamped when using physiologically plausible muscle parameters. In this paper we propose an agonist-antagonist pitch production (A2P2) model that both validates and gives insight behind the improved results of using higher-order critically damped system models in intonation modelling.

Keywords: prosody, intonation, muscle models, resonant frequency, damping

1 Introduction

Prosody is a multidimensional phenomenon comprising the intonation, energy, and duration contours of the speech signal, which carries both linguistic and paralinguistic information [3], [15]. Prosody is crucial in speech technology systems, especially in Text to Speech synthesis (TTS) where it is necessary for generating natural speech output, but also in Automatic Speech Recognition (ASR), Speech Emotion Recognition (SER) [19], emotional speech synthesis [2], and emphatic human-machine dialogue systems. Intonation is arguably the most studied and modelled dimensions of prosody [14]. Most intonation models follow one of two general approaches: (i) modelling the pitch contour directly, and (ii) modelling the underlying mechanisms, i.e. the physiology of pitch production. The physiology-based models are especially attractive because they offer insight into the way prosody is produced, and because of their inherent multilinguality.

One of the most well-known physiological models is the command-response (CR) model of Fujisaki [4], which models the pitch contour as a sum of global, phrase components, and local, accent components. Both components are output

from a 2nd order critically damped system that models laryngeal muscle activation based on the Spring-Damper-Mass (SDM) muscle model [5]. More recently, research has shown that using higher order system models increases intonation modelling performance. The quantitative target approximation (qTA) model, for example, uses a 3rd order system to generate the surface pitch contours [13]. We have also observed improved performance in our Weighted Correlation Atom Decomposition (WCAD) based intonation model[3], when higher 6th order system responses are used [9], [8]. These findings necessitate a closer examination of the muscle model used in intonation modelling.

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 [20]. Recently, we have analysed the two most commonly used muscle models: the 2nd order SDM model and the 3rd order Hill type model [7]. Research suggests that the SDM model is too simple to capture the basic mechanics of muscle activation [10]. On the other hand, the Hill type model while offering improved modelling of muscle-tendon dynamics, exhibits underdamped behaviour when using physiologically plausible muscle parameters [11]. In this paper we propose an agonist-antagonist pitch production (A2P2) model [12] and analyse how it relates to recent results in physiological intonation modelling. The analysis shows that the A2P2 model validates and gives insight behind the improved results of using higher-order critically damped system models in intonation modelling.

2 SDM and Hill muscle models

The spring-damper-mass (SDM) model shown to the left in Fig. 1 is the simplest model of muscle activation. It comprises a parallel elasticity (PE) k, a damper c and a force generator F. If we assume steady state initial conditions and an impulse driving force its transfer function is given by (1) [6]. From it, we can extract the damping ratio ζ and the undamped resonant frequency ω_0, which are given by (2). If we plug in physiologically plausible parameters taken from the elbow muscles [11] into the SDM, we obtain the zero-pole diagram and corresponding impulse responses in Fig. 2. The diagram shows that the system reaches critical damping only for $c = 10$, which is at the extreme end of the physiologically plausible range.

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

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

The three-element Hill muscle model [20] is shown in its Poynting-Thomson

[3] The WCAD implementation code is available on gitHub at `https://github.com/dipteam/wcad`

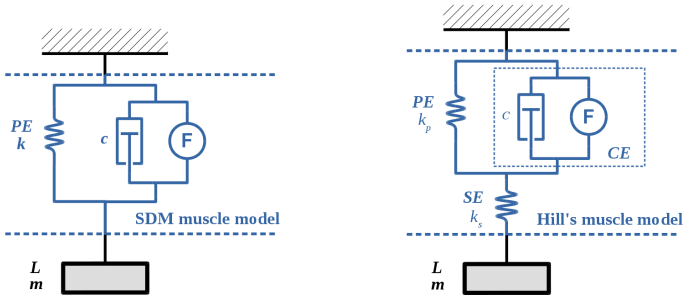

Fig. 1. The 2nd order spring-damper-mass (SDM) muscle model (left), and the Hill three element model (right).

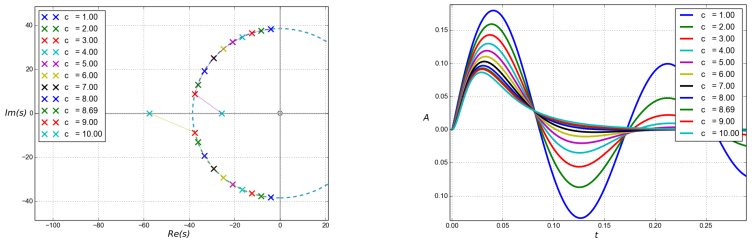

Fig. 2. Zero-pole diagram (left) and impulse response (right) of the SDM model, for a sweep of values of $c \in [1, 10]$, for $k = 178$ and $m = 0.12$.

(PT) form to the right of Fig. 1. It improves on the SDM by adding a series elasticity (SE) k_s that models the tendons connecting the muscle to the bone. It is the simplest model that takes into account the essential interactions arising from the stiffness of the tendon [10]. Its transfer function under steady state initial conditions and an impulse driving force is given in (3). To derive its resonant frequency ω_0 (4) we can use the impedance electro-mechanical analogy [1] to draw the equivalent electrical circuit, find its input impedance $Z_i(j\omega)$, and equate its imaginary part to 0 [7]. Unfortunately, there is no straightforward solution for the damping ratio ζ [11].

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

$$\omega_0^2 = \frac{k_p k_s}{m(k_p + k_s)} \tag{4}$$

The movement of the poles is shown in Fig. 3 for a sweep of muscle damping c and an increasing SE to PE ratio $k = {}^{k_s}/_{k_p}$. We can see that the system reaches critical damping only for $k \geq 8$, when the two imaginary poles reach the real axis, and is underdamped over most of the parameter range. In fact, for $k \geq 8$ the Hill model exhibits underdamped oscillatory behaviour independent of its damping c [11].

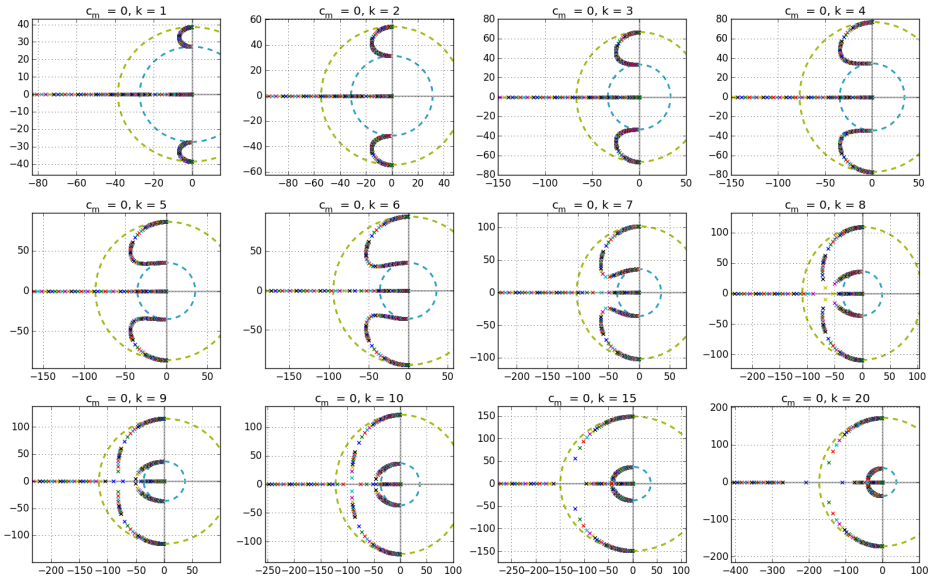

Fig. 3. Movement of the poles for the Hill model for a sweep of $c \in [0.1, 1000]$.

3 Physiology of pitch production

It is clear the the SDM, and even the Hill muscle model, with their underdamped behaviour cannot on their own account for the dynamics of the laryngeal muscle system. In order to build a better model we have to take a examine more closely the physiology of pitch production. Such a detailed analysis reveals four physiological sources of pitch change [16]:

(i) Cricothyroid (CT) muscle that rotates the thyroid cartilage in respect to the cricoid, stretching the vocal folds and raising pitch,
(ii) Vocalis (VOC) muscle, whose contraction decreases vocal cord length, but increases their tensile stress, effecting a rise in pitch [18],
(iii) Sternohyoid (SH) muscle that lowers the larynx decreasing vocal fold tension and pitch, and
(iv) Subglottal pressure (P_{SB}), which linearly correlates to pitch.

Other researchers have suggested that thyrohyoid (TH), rather than the SH muscle effectuates the drop in pitch [5], but these muscle have been found to activate in unison.

4 The agonist-antagonist pitch production model

Reflecting the complexity of the laryngeal muscle system we propose an agonist-antagonist pitch production (A2P2) model to capture the opposing muscle physiological environment of pitch production [16], [5]. The agonist-antagonist concept

was first proposed by Plamondon and colleagues [12], mainly in the context of handwriting analysis. Plamondon's model is built around the velocity of muscles following a lognormal profile. The lognormal in turn arises as a limiting case where complex muscles are driven by signals travelling some distance from the brain, and driving large masses. In considering the Hill model (and derivatives), we rather model the absolute offsets of individual muscle fibres. Of course, complex muscles lead to higher order models which likely tend towards lognormal profiles. It is an open question whether the muscles associated with prosody are small enough to be modelled as individual fibres. At least, the thrust of the present work is to understand what can be gained from assuming so. Conversely, the difference between a lognormal and the gamma-like profiles that arise from such analysis is not large, and probably below the noise level of measurements of prosody.

The A2P2 model is shown in Fig. 4 and consists of an agonist Hill muscle that models the CT and VOC muscles, an antagonist Hill muscle that models the SH-TH muscle complex, and a mass with its damper and elasticity that represents the thyroid cartilage held in place by the elasticity of the vocal folds and whose movements are damped by the friction at its joint with the cricoid. Although P_{SB} is not explicitly modelled, it is indirectly included in the two opposing muscle models, as it is also due to the activation of muscles in the respiratory system. The physiological plausibility of the proposed model is grounded on the assumption that we can group all of the muscles responsible for the production of the pitch into two equivalent opposing muscles, whilst still being small enough to merit the small muscle assumption in the Hill model. This has been common practice when modelling muscle systems [11] and is also justified by the correlation seen in the activation of the CT and the VOC [16].

Transfer function. To obtain the transfer function of the proposed model we can use the impedance electro-mechanical analogy [1] to obtain the equivalent electrical circuit shown in Fig. 5. When solving in the Laplace domain [17] we find that the proposed system is 4^{th} order with one zero. It is possible to simplify the equivalent circuit by applying Thévenin's theorem between the connection points of the two muscles, here marked A and B and thus calculating a joint equivalent Hill model for the opposing muscles. If we assume that the two opposing muscles have identical parameters, which is physiologically plausible, then the system simplifies to a 3^{rd} order system, whose impulse is given by (5) response for steady state initial conditions, and an impulsive driving force .

$$y_o(s) = \frac{1}{\frac{c_p m}{k_s} s^3 + \frac{(c_m c_p + m(k_p+k_s)}{k_s} s^2 + \frac{c_m(k_p+k_s)+c_p(k_m+k_s)}{k_s} s + \frac{k_m k_p}{k_s} + k_m + k_p} \quad (5)$$

Resonant frequency ω_0. We can now find the input impedance of the system and use it to calculate the resonant frequency ω_0. A simplified analysis, which disregards the elasticity k_m, gives (6), showing that the A2P2 model has two resonant frequencies. It is interesting to note that if we let $c_m = 0$, the simplified A2P2 model reduces to the Hill model, and as solutions of (6) we have (7).

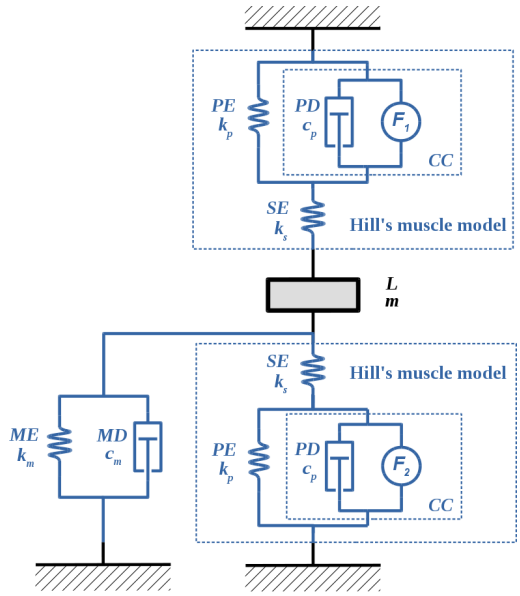

Fig. 4. The agonist-antagonist pitch production (A2P2) model.

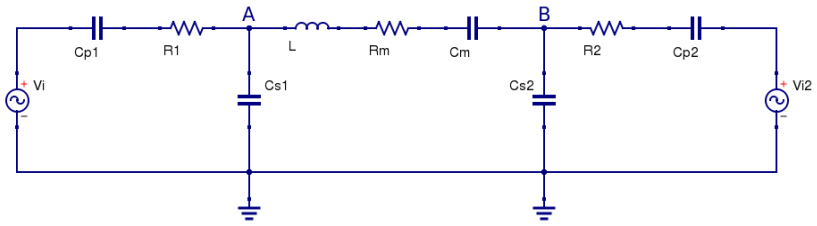

Fig. 5. Equivalent electrical circuit of the A2P2 model.

While the first solution ω_0 is equivalent to (4), ω_1 explains the outer resonant frequency seen in the movement of the poles in Fig. 3.

$$\omega_0^2 = \frac{1}{2}\left(-\frac{c_m^2}{m^2} + \frac{k_s(2k_p+k_s)}{(k_p+k_s)m} \pm \sqrt{\frac{c_m^4}{m^4} - \frac{2c_m^2 k_s(2k_p^2-3k_pk_s-k_s^2)-k_s^4 m}{m^3\,(k_p+k_s)^2}}\right) \tag{6}$$

$$\omega_0 = \sqrt{\frac{k_pk_s}{m(k_p+k_s)}} \qquad \omega_1 = \sqrt{\frac{k_pk_s+k_s^2}{m(k_p+k_s)}} \tag{7}$$

Pole movement. To understand the A2P2 model's behaviour and compare it to the Hill model, we will look at the movement of the poles in the simplified model for various k-s keeping $c_m = 2$, shown in Fig. 6. We can see that (i) for $k = 1$ the two resonant frequencies coincide, (ii) the asymptotic movement of the poles towards ω_0 and ω_1 is from the outside rather than from in between as

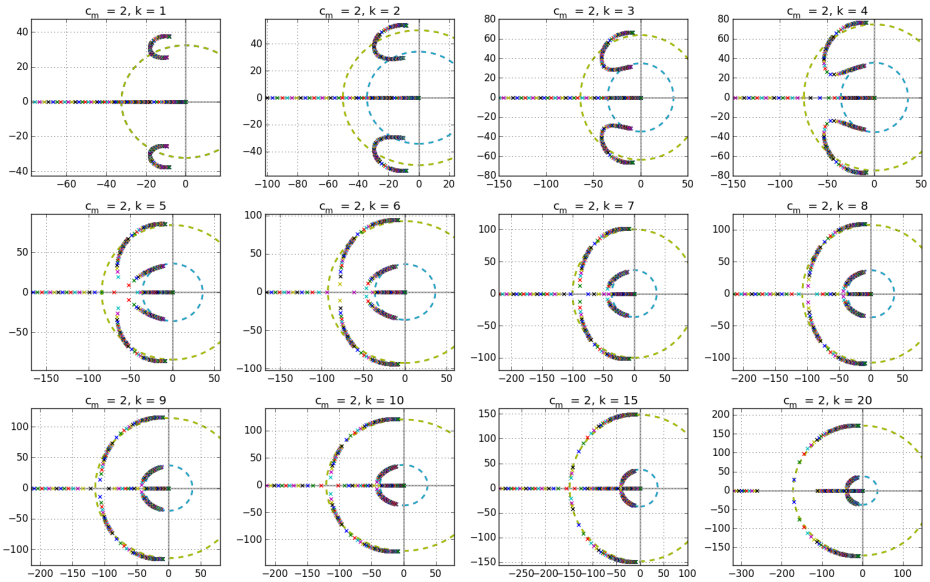

Fig. 6. Movement of the poles for the AA model for $c_m = 2$ and a sweep of $c_p \in [0.1, 1000]$.

for the Hill model, and (iii) we have critical damping already for $k = 5$, instead of $k = 8$ as was the case for the Hill model. Thus, the added damping c_m in the A2P2 compensates for the underdamped behaviour of the individual Hill model, granting critical damping and overdamping for a physiologically plausible set of parameters. This effect is emphasised if we let $c_m > 2$.

5 Conclusions

The proposed agonist-antagonist pitch production model appears to be a reasonable hypothesis for the model that is being implicitly assumed when higher orders are used in prosody models. Combined with the feedback assumption of Prom-on, it justifies use of a model order somewhere between the 2^{nd} order of the CR model and the limiting lognormal case of Plamondon. Moreover, the A2P2 model also grants physiological plausibility to the use of critically damped system models in intonation modelling.

Acknowledgements

The authors would like to thank the support of the Swiss National Science Foundation via the joint research project *"SP2: SCOPES Project on Speech Prosody"* (No. CRSII2-147611 / 1).

References

1. Beranek, L.: Acoustics sound fields and transducers. Academic Press, S.l (2012)
2. Burkhardt, F., Campbell, N.: Emotional speech synthesis. The Oxford Handbook of Affective Computing p. 286 (2014)
3. Cutler, A., Dahan, D., Van Donselaar, W.: Prosody in the comprehension of spoken language: A literature review. Language and speech 40(2), 141–201 (1997)
4. Fujisaki, H.: A model for synthesis of pitch contours of connected speech. Annual Report, Engineering Research Institute, University of Tokyo 28, 53–60 (1969)
5. Fujisaki, H.: The roles of physiology, physics and mathematics in modeling prosodic features of speech. In: Proc. of Speech Prosody (2006)
6. Garner, P.N.: A derivation of a second order damped system `http://www.idiap.ch/~pgarner/appendices/damping.pdf`
7. Gerazov, B., Garner, P.N.: An investigation of muscle models for physiologically based intonation modelling. In: Proceedings of the 23rd Telecommunications Forum. pp. 468–471. Belgrade, Serbia (November 2015)
8. Gerazov, B., Honnet, P.E., Gjoreski, A., Garner, P.N.: Weighted correlation based atom decomposition intonation modelling. In: Proceedings of Interspeech. Dresden, Germany (September 2015)
9. Honnet, P.E., Gerazov, B., Garner, P.N.: Atom decomposition-based intonation modelling. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Brisbane, Australia (April 2015)
10. Kistemaker, D.A., Rozendaal, L.A.: *In Vivo* dynamics of the musculoskeletal system cannot be adequately described using a stiffness-damping-inertia model. PLoS ONE 6(5), e19568 (05 2011), `http://dx.doi.org/10.1371%2Fjournal.pone.0019568`
11. Piovesan, D., Pierobon, A., Mussa Ivaldi, F.A.: Critical damping conditions for third order muscle models: implications for force control. J Biomech Eng 135(10), 101010 (Oct 2013)
12. Plamondon, R.: A kinematic theory of rapid human movements: Part I: Movement representation and generation. Biological Cybernetics 72(4), 295–307 (March 1995)
13. Prom-on, S., Xu, Y., Thipakorn, B.: Modeling tone and intonation in Mandarin and English as a process of target approximation. Journal of the Acoustical Society of America 125, 405–424 (January 2009)
14. van Santen, J., Mishra, T., Klabbers, E.: Prosodic processing. In: Springer Handbook of Speech Processing, pp. 471–488. Springer (2008)
15. Schuller, B., Batliner, A.: Computational paralinguistics: emotion, affect and personality in speech and language processing. John Wiley & Sons (2013)
16. Strik, H.: Physiological control and behaviour of the voice source in the production of prosody. Ph.D. thesis, Dept. of Language and Speech, Univ. of Nijmegen, Nijmegen, Netherlands (October 1994)
17. Thomas, R.E., Rosa, A.J., Toussaint, G.J.: The analysis and design of linear circuits, 7th Ed. Wiley Publishing (2012)
18. Titze, I.R., Martin, D.W.: Principles of voice production. Journal of the Acoustical Society of America 104(3), 1148 (1998)
19. Vogt, T., André, E., Wagner, J.: Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realisation. In: Affect and emotion in human-computer interaction, pp. 75–91. Springer (2008)
20. Zatsiorsky, V., Prilutsky, B.: Biomechanics of skeletal muscles. Human Kinetics (2012)