Speed-Accuracy Tradeoff In Speech Production

Venkata Praneeth Srungarapu[1], Pramit Saha[1], Sidney Fels[1]

[1]*HCT Lab, Department of Electrical and Computer Engineering, University of British Columbia*
praneethsv@ece.ubc.ca, pramit@ece.ubc.ca, ssfels@ece.ubc.ca

Abstract

This work presents our advancements in learning to solve the speed-accuracy trade-off problem in the context of speech production. We formulated a muscle-driven speech motor control task as a variant of Fitts' task. Then, we investigated whether machines can learn and demonstrate the speed-accuracy trade-off mechanism. The investigation was three-fold - first, learning to reach the target while taking into account the width of the target; second, analysis of the demonstrated behavior of an agent while varying the distance between the targets as well as the width of the targets and lastly, we plan to test the validity of Fitts' law in the case of force targets rather than kinematic targets. We present experimental results on target reaching speech motor task using a simplified 2-muscle model and perform the speed-accuracy trade-off analysis of the model based on the given task.

Keywords: speed-accuracy tradeoff, speech motor control, deep reinforcement learning

1. Introduction

Speech production is a complex biomechanical process involving the movement of vocal fold and naso-pharyngeal vocal tract (including tongue), brought about by a well co-ordinated synergy of muscle excitations. It consists of articulatory, acoustic, prosodic and communicative tasks that require precision while performing quick movements. So, speed and accuracy of such movements must have an agreement that complies with a behavioral law while performing speech motor actions. In this work, we are investigating the application of Deep Reinforcement Learning frameworks to learn an appropriate speed-accuracy relation (SAR) given a goal directed biomechanical system/task, payoff matrix and constraints, which provides significant insights into the speech production process.

2. Speech Motor Control Task

The speech motor task involves combined control of labial, jaw, tongue and vocal fold movements, which in turn is caused by careful control of the simultaneous activations of numerous interleaved muscles. It is an incredibly challenging task to control the speed-accuracy parameters of such articulatory movements by automated estimation and control of the individual muscles. In order to model such a biomechanical process, we start by simulating a two-muscle constrained system in the biomechanical toolkit, ArtiSynth (Lloyd, Stavness, and Fels 2012). We developed a system of two spring dampers conjoined by a mass to investigate the relationship between speed and accuracy as shown in Fig 1. This model is specifically worth investigating as it can be seen as a fundamental model to any speech-motor control task, including the controlled movements of vocal fold, jaw, tongue as well as other parts of the vocal tract, in the context of articulatory speech synthesis.

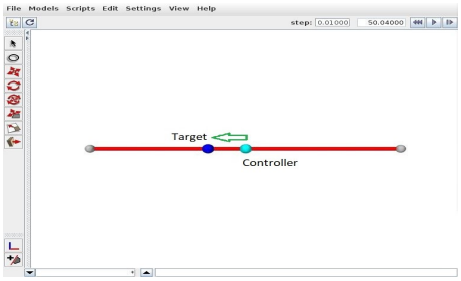


Figure 1: Our two-muscle model in a variant of Fitts' Task

3. The Proposed Methodology

In recent years, deep reinforcement learning (DeepRL) algorithms have been implemented to solve various control tasks such as gait patterns (Kidziński et al. 2018) in OpenSim simulator (Delp et al. 2007). Our primary research objective is to learn how to solve the speed-accuracy tradeoff mechanism and it is important to consider utilizing the advancements that happened in DeepRL. So, keeping in mind the speech-motor tasks, where our articulators are driven towards specific targets, we implemented a DeepRL based controller that learned how to reach a target in the kinematic space through estimation of the desired muscle activations. DeepRL utilizes neural networks as function approximators that provides optimal actions given a state-transition vector. In the spring-mass model, the action space contains muscle activations where as the state transition vector contains the position of the controller, target position as well as the target width. Based on the distance metric between the controller and target position, the agent gets an incentive; encouraging the controller to reach a target.

The spring mass model has two muscle exciters that helps the controller (shown in fig 1) to home-in on the target. The DeepRL based-controller provides the optimal muscle activations to reach the target. Depending upon the type of DeepRL method, the controller estimates the optimal muscle activations. This is particularly achieved without any prior knowledge of the system dynamics of spring-mass model and hence this method is often referred to as model-free reinforcement learning.

4. Learning to Reach

4.1. Target Reaching Task

We employed a deep reinforcement learning technique referred to as Soft Actor-Critic (SAC) (Haarnoja et al. 2018); a model-free reinforcement technique. The algorithm takes the source

and target positions as the inputs and computes the muscle activations required to move the controlled agent closer to the target point within a certain threshold.The articulators, analogous to the controlled agents mentioned here, also use this very mechanism to execute different speech-related target reaching tasks. For example: For making an oral stop consonant like /t/,/d/, /k/ or /g/, the tongue tip or body has to reach some particular target locations on the hard palate and the underlying mechanism behind such articulation can be well illustrated using the reinforcement learning technique depicted here.

4.2. Results

The algorithm achieved satisfactory results in estimating the desired muscle activations necessary to reach a target position in cartesian space. The learning performance (measured as mean reward over the episodes) of the proposed methodology is shown in Fig 2. The increase in the reward as demonstrated in the graph illustrates how successfully the algorithm estimates the optimal muscle activations necessary to reach the target, in its initial episodes.

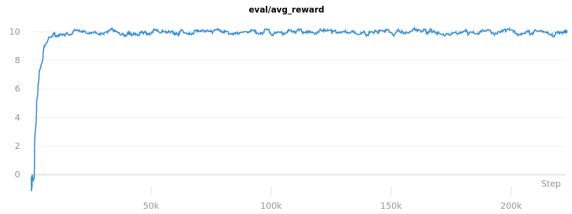


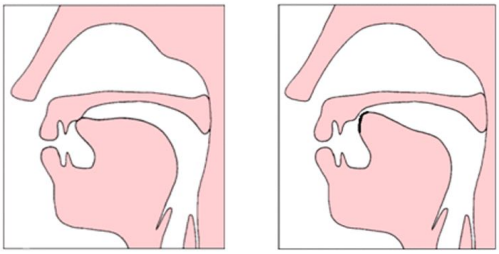
Figure 2: *Learning performance of the controller*

5. Speed-Accuracy Trade-off

Speech articulation contains complicated speech motor actions that are performed quickly, while maintaining desired accuracy. Through lots of practice, humans learn to produce rapid speech motor actions while simultaneously maintaining sufficiently high dexterity.

In order to investigate the trade-off mechanism (between the speed and accuracy of the performed tasks) through our model, we arbitrarily vary the target widths and distances during the training phase. As a result, the difficulty of the task varies accordingly, with the varying target width and distance of the target from the controller. The expected trade-off behavior is to compromise accuracy and take up the ballistic mode of movement given a large target width, where as, to compromise speed and embrace the corrective mode given a small target width.

We plan to model the speed-accuracy trade-off in such cases by leveraging the Fitts' task, *i.e.*, by means of generating targets in 1D space - on the left and right side of the controller (with a target width W and distance D within an episode). This simplified model can be considered analogous to a tongue tip reaching different positions on and near the hard palate while making continuous speech movements. For example: For the tongue to make a fricative sound is much tougher (where it needs more precision regarding placement of the tongue tip - at a particular distance vertically downwards from the hard palate) than making a stop consonant (where the tip can just strike anywhere within a wider range of positions directly on the palate) as shown in Fig 3. Similarly, reaching a distant target with smaller width is a much harder task for the controller than reaching a nearer target with much larger width. This analogy helps us to explore the Fitts' task of monitoring the speed-accuracy trade-off in the context of articulatory speech synthesis.



(a) (b)

Figure 3: Tongue positions for (a) oral stop consonants - /t/, /d/ and (b) palato-alveolar fricatives

6. Discussion and Conclusion

In conclusion, we successfully implemented a learning model to reach a target in the kinematics space. We will continue to conduct various experiments to complete our analysis of the speed-accuracy trade-off problem. One experiment worth conducting is to embed a rigid body such as wall in our spring mass model to verify the robustness of the learned trade-off behavior. Also, we would like to investigate the test case of force targets to verify whether Fitts' law holds valid in such cases.

7. Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada and Canadian Institutes for Health Research (CIHR).

8. References

Delp, Scott L., Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen (Nov. 2007). "OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement". In: *IEEE Transactions on Biomedical Engineering* 54.11, pp. 1940–1950. DOI: 10.1109/tbme.2007.901024. URL: https://doi.org/10.1109/tbme.2007.901024.

Haarnoja, Tuomas, Aurick Zhou, Pieter Abbeel, and Sergey Levine (Oct. 2018). "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. Stockholmsmässan, Stockholm Sweden: PMLR, pp. 1861–1870. URL: http://proceedings.mlr.press/v80/haarnoja18b.html.

Kidziński, Łukasz, Sharada Prasanna Mohanty, Carmichael F. Ong, Zhewei Huang, Shuchang Zhou, Anton Pechenko, Adam Stelmaszczyk, Piotr Jarosik, Mikhail Pavlov, Sergey Kolesnikov, Sergey Plis, Zhibo Chen, Zhizheng Zhang, Jiale Chen, Jun Shi, Zhuobin Zheng, Chun Yuan, Zhihui Lin, Henryk Michalewski, Piotr Milos, Blazej Osinski, Andrew Melnik, Malte Schilling, Helge Ritter, Sean F. Carroll, Jennifer Hicks, Sergey Levine, Marcel Salathé, and Scott Delp (2018). "Learning to Run Challenge Solutions: Adapting Reinforcement Learning Methods for Neuromusculoskeletal Environments". In: *The NIPS '17 Competition: Building Intelligent Systems*.

Lloyd, John E, Ian Stavness, and Sidney Fels (2012). "ARTISYNTH: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation". In: *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. Springer, pp. 355–394.