TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.6, June 2014, pp. 4115 ~ 4124
DOI: 10.11591/telkomnika.v12i6.5388

Constructing Cerebellum Model by Researching on its Contributions to DIVA

Yuanyuan Wu*[1], Shaobai Zhang[2]
College of Computer, Nanjing University of Posts and Telecommunications
No.66 Xinmofan Road, Nanjing, Jiangsu
*Corresponding author, e-mail: 525015923@qq.com[1], adzsb@163.com[2]

Abstract

DIVA (Directions into Velocities of Articulators) is a mathematical model of the processes behind speech acquisition and production, supposed to achieve a functional representation of areas in the brain that are involved in speech production and speech perception. Introducing cerebellum control mechanism into the model plays a significant role in improving the mechanism of speech acquisition and production based on DIVA model. The paper studies its learning process, and explores cerebellar contributions to the model, that is feedforward learning, sensory predictions, feedback command production and the timing of delays, and then constructs a cerebellum model that is closer to neuroanatomy and is applied to DIVA model. Simulation results show that the improved DIVA model can produce more clear and explicit speech sounds, and is more close to human-like pronunciation system. The cerebellum model that established in this paper can be applied to speech acquisition and production based on DIVA model.

Keywords*: cerebellar roles, cerebellum model, DIVA model, speech acquisition and production, motor learning*

1. Introduction

DIVA model was first proposed by Guenther in 1994 [1], and then has been improved a lot till now. The earlier versions of DIVA models have some disadvantages more or less [2]. In order to solve these problems, a DIVA model [5] that is closer to neuroanatomy is proposed by Guenther and Ghosh. The model uses double-sensory, auditory and somatosensory, as the benchmark structure, and defines the components that are involved in premotor area, motor area and auditory and somatosensory areas in cerebral and cerebellar cortex, which establishes a corresponding relationship between the components and actual neuroanatomy. Moreover, the model combines feedforward control subsystem with feedback control subsystem to control articulator movements that contain realistic neural processing delays, and computer simulations of the model are presented to illustrate that the model can provide a detailed account for experiments involving compensations to perturbations of the lip and jaw. Although a lot of improvements have been done on DIVA model, some other performance factors are not taken account of, and cerebellar learning mechanism is not integrated into the model to account for the timing of delays and sensory motor learning in neural transmission.

The cerebellum has an anatomical structure that is consistent throughout. Because of its unique internal structure and the widespread connectivity to and from cerebral cortex, several hypotheses that the cerebellum utilizes a consistent processing scheme of transforming inputs to outputs have been proposed. Allen and Tsukahara first proposed the cerebrocerebellar interaction theory in 1974 [8]. And in 1998, Miall et al. proposed the concept of internal model [10]. They believed that the cerebellum contains two varieties of internal model, forward and inverse models. The forward model predicts the consequence of a motion or an action, and the inverse model provides the essential commands to accomplish the motion or action. Meanwhile, the cerebellum may contribute as a delay model that queues sensory predictions to match with actual sensory feedback. These theories above have laid the foundation for the integration of cerebellar roles into motor control and learning system.

Although DIVA model is used to represent the function of areas in the brain that are involved in speech production and speech perception, these studies show that the cerebellum is also an indispensable part of the model, and the global cerebrocerebellar circuitry has been well

Received December 29, 2013; Revised March 3, 2014; Accepted March 18, 2014

established. The integration of the cerebellum into DIVA model plays a significant role in perfecting the mechanism of speech acquisition and production. However, how to integrate the cerebellum into this mechanism, and which roles does the cerebellum play in the whole processing? In this paper, these problems will be discussed.

2. DIVA Model and its Learning Process

DIVA model (see Figure 1) is a mathematical model that describes the processes of speech acquisition and production, and is used to represent the function of areas in the brain that are involved in speech production and speech perception. The model is an adaptive neural network that learns to control movements of simulated speech articulators in order to produce words, syllables, or phonemes [6]. It consists of integrated feedforward and feedback control subsystem. It takes a speech sound string as input to generate a time sequence of articulator positions that command movements of the simulated vocal tract. Figure 1 is the current DIVA model block diagram.

Before DIVA model can produce speech sounds, the mappings between each component of the model must be learned. In order to investigate cerebellar contributions to DIVA model, the first thing is to explicit the learning process of the mappings between the various components of the model. The whole learning process is divided into two phases, early babbling phase and imitation phase. Figure 2 is a simplified DIVA model block diagram which indicates the mappings tuned during the two learning phases. During a babbling phase, somatosensory and auditory feedback signals, which are used to learn the mappings between different neural representations, are provided by random movements of the speech articulators. After babbling phase, the model goes into imitation phase in which the model can quickly learn to produce either new sounds from audio samples provided to it or arbitrary combinations of the sounds it has learned.

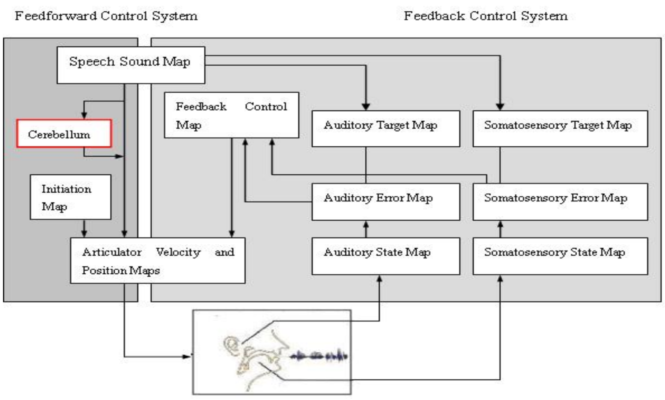


Figure 1. The DIVA Model Block Diagram

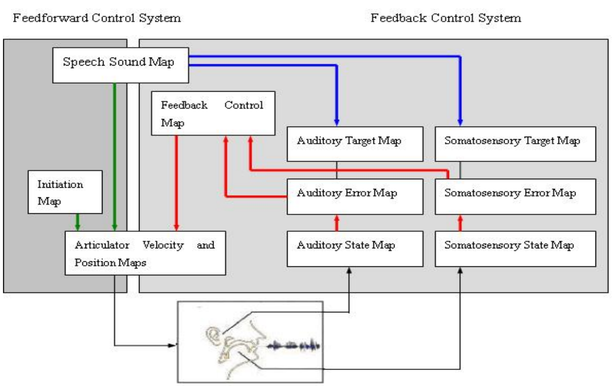


Figure 2. Learning in the DIVA model

2.1. Early Babbling Phase

During a process similar to infant babbling, the model first learns the relationship between motor commands and corresponding sensory outcomes. Somatosensory and auditory feedback that causes motor commands is provided by pseudo-random articulator movements. The motor commands and their sensory consequences are used to tune the synaptic projections from sensory error maps to the feedforward control map (red arrows in Figure 2). Once tuned, these projections transform sensory error signals into corrective motor velocity commands.

2.2. Imitation Phase

After babbling phase in which the general sensory-to-motor mapping has been learned, the model goes into a second learning phase, imitation phase, to produce speech sounds. There are two kinds of mappings need to be learned in this phase, one is the mapping in the feedback control system from speech sound map to auditory and somatosensory target map (blue arrows in Figure 2), and the other is in the feedforward control system from speech sound map to articulator velocity and position maps (green arrows in Figure 2). In the former, analogous to the sounds of the native language of an infant, the model is presented speech sound samples which take the form of time varying acoustic signals spoken by a human speaker. Once a new speech sound is presented, it becomes associated with an unused cell in speech sound map. Subsequently, the model learns an auditory target for that speech sound in the form of a time-varying region. In this way, the corresponding relationship between the cell in speech sound map and the auditory target in auditory target map is established, that is to say, weights from speech sound map to auditory target map are tuned. In addition, weights from speech sound map to somatosensory target map are tuned during correct self productions.

Once auditory targets have been learned, the second kind of mapping are also learned during the imitation phase. Because that the projections from speech sound map to articulator velocity and position maps are tuned poorly, and production relies heavily on the feedback control system, large sensory error signals are produced in the initial attempts to produce the speech sound. However, the feedback-based corrective motor command in each production is added to the weights from speech sound map to articulator velocity and position maps, incrementally improving the accuracy of the feedforward motor command. With practice, the feedforward commands are able to produce the speech sound with minimal sensory error. Therefore, unless unexpected sensory feedback is encountered, the production of the speech sound little relies on the feedback control system.

3. The Proposed Method

Based on the cerebellar activity that is noted by neuroimaging studies of motor learning, Guenther put the cerebellum control module into the projections from speech sound map to articulator velocity and position maps in feedforward control system of the DIVA model (the cerebellum module highlighted by a red outline in Figure 1), in order to learn and maintain feedforward motor commands. On the basis of the anatomical structure and observed neurophysiology, several functional roles have been hypothesized for the cerebellum, including tonic reinforcement, timing of behavior, command-feedback comparison, combining and coordinating movements, sensory processing and motor learning [8]. Therefore, the cerebellum control module can be applied to not only feedforward control system of the DIVA model but also feedback control system.

3.1. Adding the Cerebellum Module to the Projection from Feedback Control Map to Articulator Velocity and Position Maps

Kawato and colleagues have proposed a cerebellar feedback-error learning model [10], as shown in Figure 3. The controlled object in Figure 3 is a physical entity that needs to be controlled by the central nervous system (CNS), such as the eyes, hands or legs. It can be considered as a cascade of transformations between motor command and linkage motion, and between this linkage motion and the controlled object motion. The inverse model is considered as a neural representation of the transformation from the desired movement trajectory of the controlled object to the corresponding motor commands. The feedback controller converts trajectory error into a corrective feedback motor command which is used as a teaching signal to train the inverse model. Because the transfer characteristics of the inverse model are the inverse of those of the controlled object, the cascade of the two systems gives an approximate identity function. That is, if a desired trajectory is given to the inverse model, then at the end of the cascade, the actual trajectory will be fairly close to the desired trajectory. Thus, the accurate inverse model can be used as an ideal feedforward controller, and its output signal is called feedforward motor command.

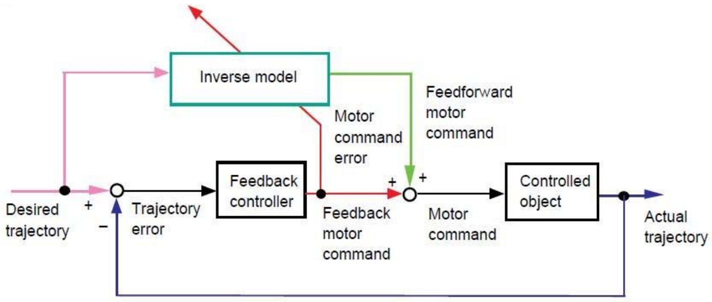


Figure 3. The Cerebellar Feedback-error Learning Model

The cerebellar feedback-error learning model above also can be applied to speech acquisition and production based on DIVA model (see Figure 4). The desired sensory targets in auditory and somatosensory target maps are compared to the current auditory and somatosensory states, and the error signal arises. The error signal is then mapped into appropriate corrective motor command via feedback control maps. The feedback motor command, on the one hand, is used as a teaching signal, with the desired sensory target trajectory which is used as a contextual signal, to train the inverse model of the cerebellum. Thus, the corresponding feedforward motor command is learned to produce. On the other hand, the feedback motor command is integrated and combined with the feedforward motor command in articulator velocity and position maps to control muscles of the face and vocal tract to produce the speech sound. The actual auditory and somatosensory states of the current speech sound are applied to the next circulation again. As a result, the cerebellar feedback-error learning model can be added to DIVA model. The corrective feedback motor commamd is received by the cerebellum and used as a teaching signal to train the inverse model to learn to produce feedforward motor command. From the above we know the cerebellum is not only involved with the learning and maintenance of feedforward motor commands but also receives the corrective feedback motor command as a teaching signal.

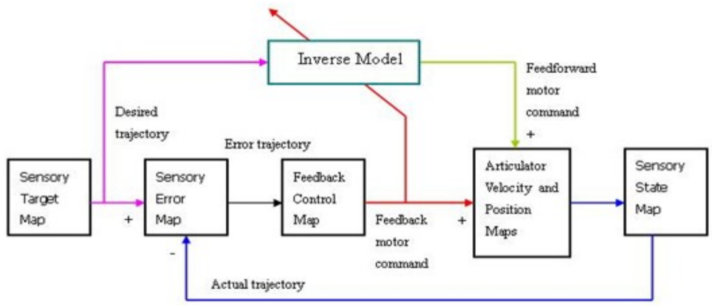


Figure 4. The Cerebellar Feedback-error Learning Model Based on DIVA Model

In addition, neuroimaging studies of motor learning have noted cerebellar activity that is associated with the size or frequency of sensory error. It is hypothesized that the cerebellum makes a contribution to the feedback motor command, and a representation of sensory errors in the cerebellum drives corrective motor commands and contributes to feedback-based motor learning. The functional role of the cerebellum is in accord with that of the projection from feedback control map to articulator velocity and position maps. The current auditory and somatosensory states which are available through sensory feedback are compared to these targets in the higher-order auditory and somatosensory cortices. If the current sensory state falls outside of the target region, an error signal arises. These error signals are transmitted to feedback control map, and then mapped into appropriate corrective motor commands via learned projections from the sensory error cells to the motor cortex. This mapping from desired sensory outcome to the appropriate motor action is an inverse kinematic transformation and is the functional role of the inverse model of the cerebellum. As a result, we add the cerebellum module to the projection from feedback control map to articulator velocity and position maps. The added cerebellum module is highlighted by a green outline in Figure 5.

3.2. Adding the Cerebellum Modules to the Projections from Speech Sound Map to Auditory and Somatosensory Target Maps

The forward model provides the crucial state estimates that can predict the outcome of motor action. For example, in visually guided tracking tasks, the subject tries to control the position of his or her hand via visual information from the target and the hand. This information is delayed due to visual processing and does not directly inform the CNS about the changes in muscle forces or even joint angles to correct any motor errors. However, a forward model can provide the missing feedback information and solve the problem.

In DIVA model, the projections from speech sound map to sensory target map predict the sound of the speaker's own voice while producing the sound based on auditory examples from other speakers producing the sound, as well as one's own previous correct productions. The cerebellum uses sensory error to build forward models to generate sensory predictions. Therefore, the cerebellum contributes to the attenuation of sensory target representation in sensory cortex.

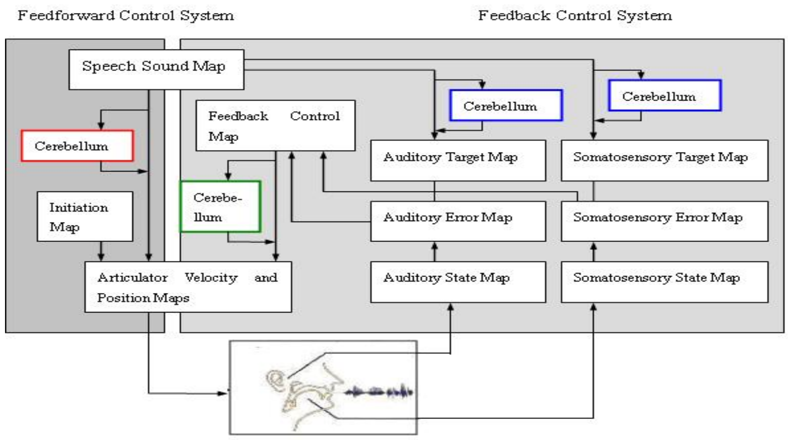


Figure 5. The DIVA Model that Adds the Cerebellum Control Modules

Moreover, DIVA model contains not only intrinsically cortico-cortical delays but also a kind of learned or necessary timing of delays. For example, the delays between premotor cortex and motor cortex are set to make the learning signals arrive at the motor cortex at the same time as the corresponding feedback corrective command signal, so that the correct portion of the feedforward command are adapted. The delays between premotor cortex and auditory/somatosensory areas are also set to make the auditory/somatosensory expectation signals arrive at the error maps at the same time that the corresponding auditory/somatosensory state signals, so that the error signals are computed correctly. Some studies have proposed the cerebellum can be used as a delay model that queues the sensory predictions to match with actual sensory feedback. The cerebellum delays signals for an appropriate duration or triggers appropriate parts of cerebral cortex at the proper times and as a locus for learning feedback commands. Hence, the cerebellum is a likely contributor to deal with the timing of delays.

On the basis of the cerebellar functional roles above, we add the cerebellum module to the projection from speech sound map to auditory target map, the same as to somatosensory target map. That is, this mapping may include a trans-cerebellar contribution in addition to a cortico-cortical contribution. The added cerebellum modules are highlighted by blue outlines in Figure 5.

4. Research Method

4.1. Structure of the Cerebellum Model

The cerebellum model that applies to DIVA model in this paper is configured based on neuroanatomy [13]. Figure 6 shows the structure of the cerebellum model which is formed by 120 granular (Gr) cells, 1 Golgi (Go) cell, 6 basket and stellate (Ba/St) cells, and 1 Purkinje (Pk) cell. The number of each cell is on the basis of the actual ratio of the cell population as much as possible [13]. Mossy fibers (mf) [15] deliver the inputs that carry a desired trajectory to Go cells and Gr cells. Go cells receive excitatory input from Gr cells as well, and simultaneously inhibit Gr cells, forming a negative feedback loop. The excitatory outputs of Gr are also received by Pk cells and Ba/St cells. Meanwhile, Ba/St cells inhibit Pk cells, forming a negative feedforward pathway. The Pk cells are the sole output of the cerebellum and a significant part of this output reaches cerebral cortex via the thalamus. In addition, a climbing fiber (cf) delivers another input that carries a control error signal to each Pk cell. By adjusting the synaptic efficacies between Gr and Pk, the output of a Pk cell can be modified to reduce the error signal. So when Gr and cf are both activated, the synaptic efficacies decrease, forming long term depression (LTD), whereas when Gr is active alone the synaptic efficacies increase, forming long term potentiation (LTP) [16]. Synaptic weights for these connections above were either positive or negative random numbers depending on the type of synapse, that is, excitatory or inhibitory. And only the synapses between Gr cells and a Pk cell can be modifiable.

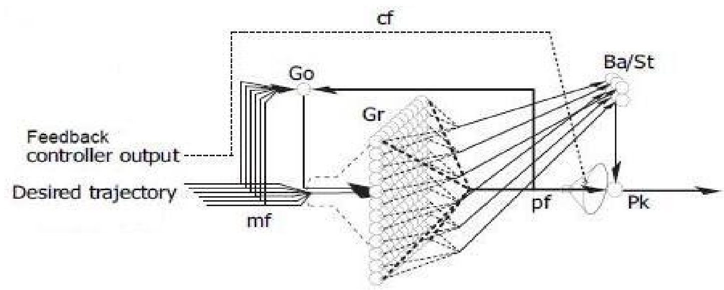


Figure 6. The Structure of the Cerebellum Model

Each cell type is described as follows, where Y is the output of each cell and W_{ji} is the synaptic weight between a cell i and a cell j.

Golgi cell:

$$X_{Go} = \sum_{mf=1}^{6} W_{Gomf} \cdot Y_{mf} + \sum_{Gr=1}^{120} W_{GoGr} \cdot Y_{Gr} \quad (1)$$

$$Y_{Go} = \frac{2}{1 + e^{-\frac{X_{Go}}{\mu}}} - 1 \quad (2)$$

Granule cell:

$$X_{Gr} = \sum_{mf=1}^{6} W_{Grmf} \cdot Y_{mf} + W_{GrGo} \cdot Y_{Go} \tag{3}$$

$$Y_{Gr} = \frac{2}{1+e^{-\frac{X_{Gr}}{\mu}}} - 1 \tag{4}$$

Basket cell:

$$X_{Ba} = \sum_{Gr=1}^{20} W_{BaGr} \cdot Y_{Gr} \tag{5}$$

$$Y_{Ba} = \frac{2}{1+e^{-\frac{X_{Ba}}{\mu}}} - 1 \tag{6}$$

Purkinje cell:

$$Y_{Pk} = \sum_{Gr=1}^{120} W_{PkGr} \cdot Y_{Gr} + \sum_{Ba=1}^{6} W_{PkBa} \cdot Y_{Ba} \tag{7}$$

4.2. Learning Algorithm

This paper employs the feedback-error learning scheme [17] for the learning algorithm of the cerebellum model. The synaptic efficacies between Gr and Pk cells are modified to minimize the output of the feedback controller by implementing the following equations, where $W_{PkGr}(t)$ is a synaptic weight between a Gr cell and a Pk cell at time t, γ is a learning rate, and E_{cf} is the activity of cf input.

$$\Delta W_{PkGr}(t) = \gamma \cdot Y_{Gr} \cdot E_{cf} \tag{8}$$

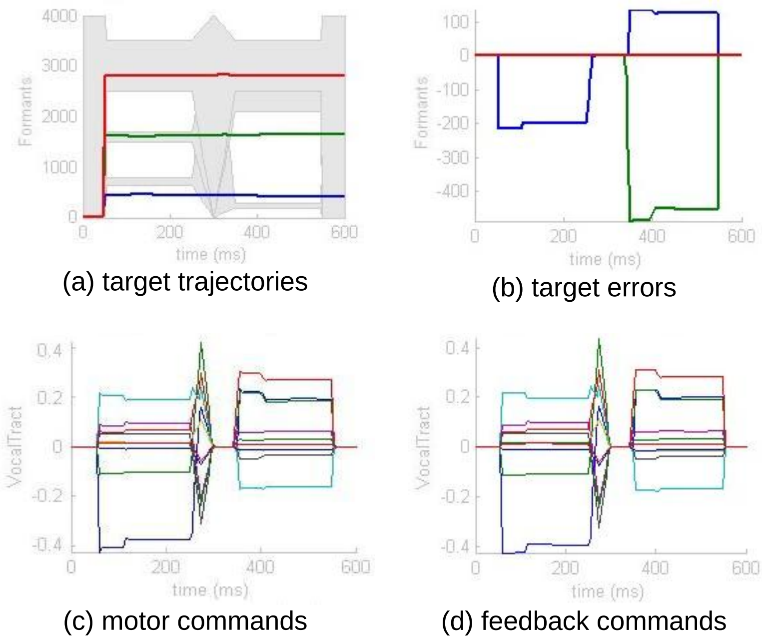
$$W_{PkGr}(t) = W_{PkGr}(t-1) + \Delta W_{PkGr}(t) \tag{9}$$

5. Results and Analysis

For the experimental simulation, we gradually add the corresponding cerebellum module to the subsystem on the basis of feedback-based DIVA model, and compare differences of formant frequencies and articulator positions of DIVA model that is added before and after when producing the utterance /adi/.

Figure 7 shows the formant frequencies and articulator positions of feedback-based DIVA model that contains none of cerebellar modules when producing the utterance /adi/. From Figure 7 (a), (b) we can see, the formants of target trajectories fall outside of the expected region and big errors of the second and third formant frequencies appear. Besides, by comparing the articulator positions of motor commands with that of feedback commands that Figure 7 (c), (d) show we know, both are almost the same. That is, the motor commands that control the articulator positions depend entirely on the feedback commands, and the feedforward commands can be learned.

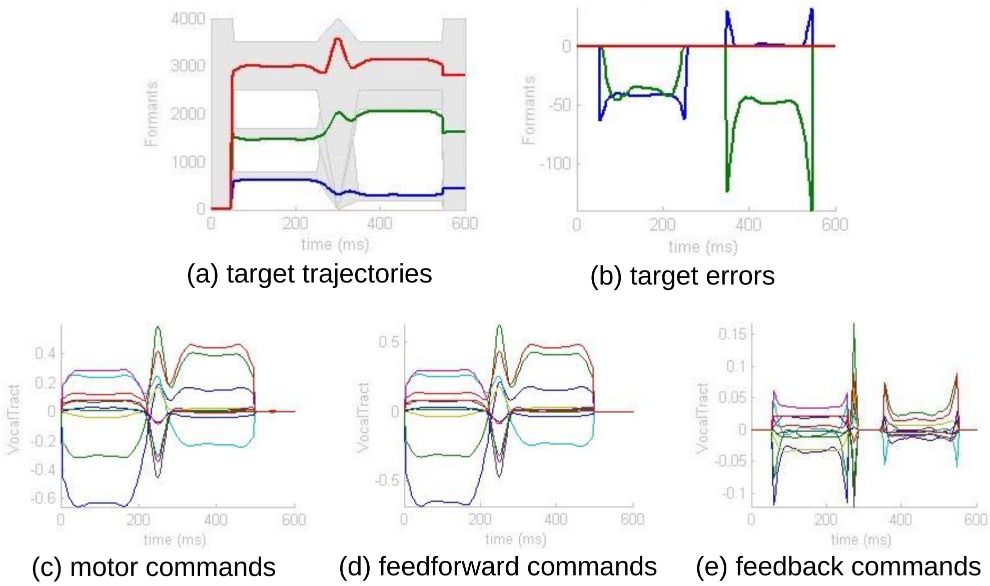
First of all, we add the cerebellum module to the projection from speech sound map to articulator velocity and position maps in feedforward system. The experimental result of simulation is shown in Figure 8. Figure 8 (a) (b) indicate the formants of target trajectories fall within the expected region and the errors of the second and third formant frequencies reduce substantially. From Figure 8 (c) (d) (e) we can see, the articulator positions of motor commands and feedforward commands are almost the same. That is, feedforward commands are integrated and combined with feedback commands to control articulator in every attempt to produce the speech sound. With practice, the feedforward commands are able to produce the speech sound with minimal sensory error. Therefore, unless unexpected sensory feedback is encountered, the production of the speech sound little relies on the feedback control system. Thus, it plays a crucial role to add the cerebellar module into the feedforward system for the feedforward command learning. We predict people with cerebellar damage may have difficulty in learning new sounds.



(a) target trajectories (b) target errors

(c) motor commands (d) feedback commands

Figure 7. The Formant Frequencies and Articulator Positions of Feedback-based DIVA Model that Contains None of Cerebellum Modules when Producing the Utterance /adi/

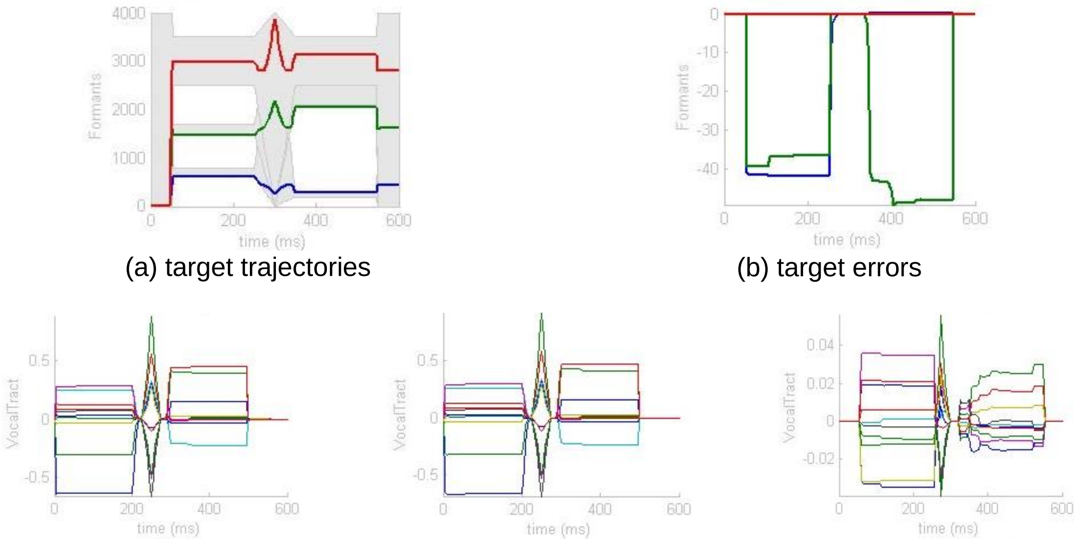


(a) target trajectories (b) target errors

(c) motor commands (d) feedforward commands (e) feedback commands

Figure 8. The Formant Frequencies and Articulator Positions of DIVA Model that is Added Cerebellum Module in Feedforward System when Producing the Utterance /adi/

Next, on the basis of the above model, we add the cerebellum modules to the projections from speech sound map to auditory/somatosensory target map and from feedback control map to articulator velocity and position maps in feedback system. Figure 9 shows the experimental result. Figure 9 (a). (b) indicate the formants of target trajectories fluctuate over an expected region more smoothly and steadily with smaller and negligible errors. The articulator positions of motor commands in Figure 9 (c), (d). (e) locate more clear and produce smaller fluctuations than that in Figure 8. Moreover, adding cerebellar modules diminishes the attempts to produce the speech sound, and accelerates feedforward commands learning process.



(a) target trajectories (b) target errors

(c) motor commands (d) feedforward commands (e) feedback commands

Figure 9. The Formant Frequencies and Articulator Positions of DIVA Model that is Added Cerebellum Module in Feedback System when Producing the Utterance /adi/

6. Conclusion

In order to improve mechanisms of speech acquisition and production based on DIVA model, and make robots have more human-like pronunciation system by using the improved model, we explore cerebellar contributions to DIVA model, such as feedforward learning, sensory predictions, feedback command production and the timing of delays, and construct a cerebellum model that is closer to neuroanatomy, and then add it to the current DIVA model. Simulation results show the cerebellum model that established in this paper can be applied to speech acquisition and production based on DIVA model. However, there are several issues about the learning process that need to be resolved. First, it remains to be determined how much of the learning of the feedforward command is transferred from the cerebellum to premotor cortex. Second, if the cerebellar circuit is necessary in learning the feedforward command, the model would predict that people with cerebellar damage may have difficulty in learning new sounds.

Acknowledgements

The research work was supported by National Natural Science Foundation of China under Grant No. 61073115 and No.61373065.

References

[1] Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. *Biological cybernetics*. 1994; 72: 43-53.
[2] Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*. 1995; 102: 594-621.

[3] Guenther FH. *A theoretical framework for speech acquisition and production.* In Proceedings of the Second International Conference on Cognitive and Neural Systems. Boston: Boston University Center for Adaptive Systems. 1998.
[4] Guenther FH, Ghosh SS, Nieto-Castanon A. *A neural model of speech production*. In Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia. 2003; 85-90.
[5] Ghosh SS. Understanding cortical contributions to speech production through modeling and functional imaging, Doctoral dissertation: Boston University Thesis, 2005.
[6] Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*. 2006; 96(3): 280-301.
[7] Tourville JA, Guenther FH. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*. 2011; 26(7): 952-981.
[8] Allen GI, Tsukahara N. Cerebrocerebellar communication systems. *Physiological review*. 1974; 54(4): 957-1006.
[9] Bastian AJ, Thach WT. Structure and function of the cerebellum. In Manto, M. and Pandolfo, M., editors, The cerebellum and its disorders. Cambridge University Press. 2001: 49-68
[10] Wolpert D, Miall C, Kawato M. Internal models in the cerebellum. *Trends in Cognitive Science*. 1998; 2(9): 338-347.
[11] Attwell PJ, Ivarsson M, Millar L, Yeo CH. Cerebellar mechanisms in eyeblink conditioning. *Annals of the New York Academy of Sciences*. 2002; 978: 79-92.
[12] Bastian AJ, Thach WT. Structure and function of the cerebellum. In Manto, M. and Pandolfo, M., editors, The cerebellum and its disorders.Cambridge University Press. 2001.
[13] Cajal S. New ideas on the structure of the nervous system in man and vertebrates. MIT Press. 1990.
[14] Barlow JS. The cerebellum and adaptive control. Cambridge University Press. 2002.
[15] Shinoda Y, Sugiuchi Y, Futami T, Izawa R. Axon collaterals of mossy fibers from the pontine nucleus in the cerebellar dentate nucleus. *Journal of neurophysiology*. 1992; 67(3): 547-560.
[16] M Ito. *Cerebellar control of the vestibule-ocular reflex-around the flocculus hypothesis*. Ann. Rev. Neurosci., 1982; 5: 275–296.
[17] H Gomi, M Kawato. Adaptive feedback control models of the vestibulocerebellum and spinocerebellum, *Biological Cybernetics*. 1992; 68: 105-114.