http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (20): 5896-5900, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.5896.5900
© 2013 Asian Network for Scientific Information

Research and Application of Constant Delay Based on DIVA Model

Feng Song and Zhang Shaobai
School of Computer Science and Technology, Nanjing University of Posts and Telecommunications,
Jiang Su, 210003, Nanjing, China

Abstract: DIVA (Direction Into Velocities of Articulators) model can be more accurately described as a mathematical model, it shows the role of the human brain regions involved in speech production and understanding. With high accuracy and openness, this model in recent years has been studied by various concerns. But we find the simulation have some access with realistic physiological processes because that the voice processing is too complex. Thus, this study uses another delay mechanism, which is constant delay mechanism, makes its workflow more reasonable and having a realistic sense of simulation results. In the simulation experiment, after introducing the new delay module, the speech process is satisfactory and closer to the physiological processes.

Key words: DIVA model, delay, constant delay mechanism

INTRODUCTION

Speech acquisition and production is a complex cognitive process involves many brain regions, so the neural network model needs to consider the interaction of various sensory and brain regions (Chua and Yang, 1988). Requirements of speech acquisition and production unified computing model promoted the emergence and development of DIVA model. Through the establishment of corresponding to each module of speech and comprehension of related brain regions, DIVA model can successfully describe and simulate the speech process (Beal, 2011).

DIVA MODEL

Introduction of the DIVA model: At present, DIVA model is the only model using pseudo-inverse control scheme and provides a precise exposition for human articulator kinematics data and related brain areas working mechanism (Tourville and Guenther, 2011).

Different DIVA models basically reflect the relationship between neuroanatomy and brain regions. They mainly define the sound channel model, the position and direction vector of the vocal organs, planning position vector and the planning direction vector, the voice sounding neuron group, mapping learning mechanism, control mechanism (Golfinopoulos *et al*., 2011).

Working mechanism of each module in DIVA model: The DIVA model is mainly composed of Speech Sound Map (SSM), Auditory State and Error Maps, Somatosensory State and Error Map and Articulatory Velocity and Position Map, respectively corresponding to four brain regions (Guenther, 2008).

Speech sound map (left ventral premotor map): Inputs of SSM module are from a list of production neurons, once any of them is activated, the module will select a group of related special neurons through a set of synaptic weights (Guenther *et al*., 2006). These neurons sparsely code a temporal representation for each production. When SSM detects any activation of GO signal, it generates a sweep of activation through the last set of activated neurons. The activation of these neurons is delivered to different areas through various delayed channels:

- Auditory and Somatosensory Error Maps, where activations were delayed to coincide with the arrival in the area of auditory signals and abstract concepts, reactions associated with this product
- Two outputs to Motor Cortex Articulatory velocity and position Maps, one without further delay and a second signal through the cerebellum implementing a learning delay to coincide with the arrival in Motor Cortex of the auditory and somatosensory compensatory feedback motor command signal associated with this production

---

Corresponding Author: Feng Song, School of Computer Science and Technology,
Nanjing University of Posts and Telecommunications,
Jiang Su, 210003, Nanjing, China

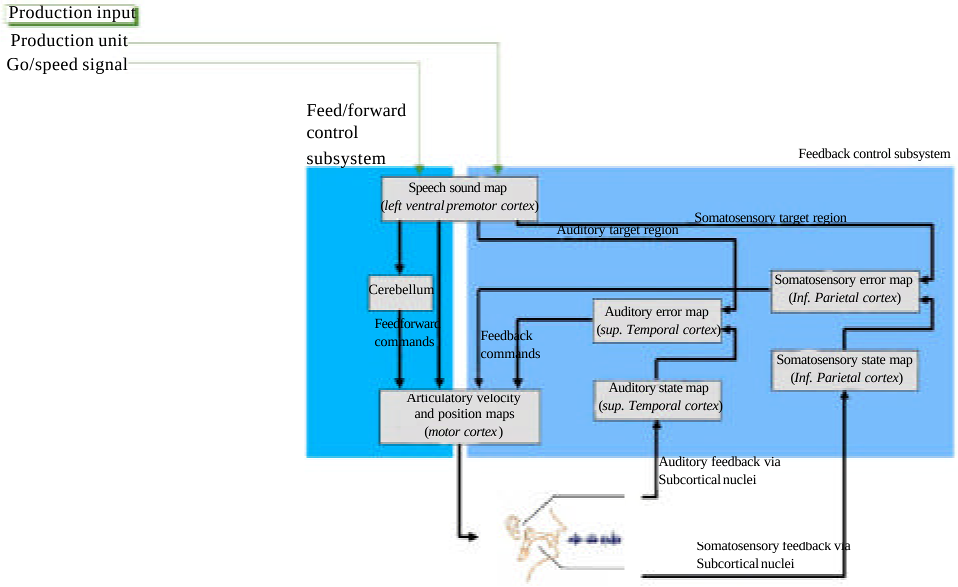

Fig. 1: DIVA model

Auditory state and error map (superior temporal cortex): Through subcortical nuclei, auditory state information reaches the superior temporal cortex after a fixed delay to activate a group of auditory neurons encoding the current state (Brumberg *et al*., 2010). These neurons encode pitch and formant position which characterize the current vocal state.

Outputs from SSM and to superior temporal cortex are converted to a desired auditory target representation through a set of synaptic weights. Target regions are defined by two separate groups of neurons, encoding respectively the minimum and maximum values for each auditory feature. These signals are then compared with activations in the Auditory state maps to generate error signal, representing the distance between the current state and target region, which is then projected to the right PreMotor Cortex.

Somatosensory state and error map (inferior parietal cortex): Somatosensory information is processed in Somatosensory state and error maps and follows the same steps in the Auditory state and error maps,but the delays associated with the somatosensory feedback signal are typically lower than those associated with auditory feedback signals. Somatosensory neurons sparsely encode the vocal tract state including the place of articulation, the glottal pressure and the vocal fold oscillation information. The information reaches Inferior Parietal Cortex from subcortical nuclei, where the somatosensory error signal is computed and projected to right ventral preMotor cortex.

Articulatory Velocity and Position Map (Motor Cortex): After implementing an inverse map of the motor direction to auditory and somatosensory direction transformation for each possible motor state, inputs from Auditory and Somatosensory error areas are converted to corrective motor commands throuth a set of learned weights. Than these motor correction signals are combined to form a Feedback motor command (Brown *et al*., 2008). Inputs from SSM are converted to a corresponding feedforward motor command through a set of synaptic weights. After a small decay along the null-space direction of the previous articulatory position, these feedback and feedforward signals are then combined to define the current articulatory position which can control the vocal tract.

As a learning signal for the feedforward weights, the auditory and somatosensory feedback motor command signals are also tranced to Motor Cortex and combined with the learning signal from Cerebellum.

The current articulatory position is also projected via a learned delay to inform the inverse map computations on the articulatory state at the time of the error-generating production.

Signals from SSM are combined via a learned set of auditory and somatosensory delays and then projected to Articulatory Velocity and Position Maps.

Elements of DIVA model: Each module in the DIVA model consists of several basic elements; some of them are provided by MATLAB, such as the adder components, maximum and minimum value calculator components, while the others by DIVA model itself, the most important are these six elements: Delay elements, gain elements, slider-gain elements, synaptic weight elements, inverse-map elements, Null-space projector element.

Delay elements: Small green boxes in a Simulink diagram are corresponding to fixed-delay elements. The Simulink DIVA model is implemented using discrete fixed-step elements, where a unit delay corresponds to the fixed-step size and currently 5 m is the fundamental sample time.

Gain elements: A small triangular icon in a Simulink diagram represents a gain factor in the model.

Slider-gain elements: Small black boxes in a Simulink diagram correspond to the activations of neural populations of interest. Synaptic weight elements: There are two types of synaptic weight elements in the Simulink DIVA model:

- The standard weight element, it implements weights that can be modeled as a simple matrix multiplication
- The adaptive weight element, it is just a standard weight element where the weights can be adaptively modified through additional learning inputs

Inverse-map elements: These elements implement an clear computation of the inverse Jacobian of the vocal tract model and when locate in the Motor Cortex module they implement the auditory and somatosensory inverse maps. Users are allowed to modify:

- The portion of the output of the vocal tract model to be used in each case
- The Jacobian estimation step size
- the Jacobian regularization factor

Null-space projector element: This element locates in the Motor Cortex module and implements a null-space projection of the current articulatory state. The null-space projection is defined as the operation that projects an input vector on the null-space of the Jacobian of the forward map at the current articulatory state. Users are allowed to define:

- A multiplicative gain factor which is used to be applied to the null-space projection output
- The portion of the output of the vocal tract model to be used when estimating the forward map null-space
- The Jacobian estimation step size
- The Jacobian regularization factor

CONSTANT DELAY BASED ON DIVA MODEL

Constant delay: Chua and Roskaintroduced the delayed cellular neural networks in 1990, then many scholars devote themselves to the study of it (Zhang, 2012). Most people would use the Lyapunov method combined with some inequality analysis in order to obtain ideal results in the discussion of this kind of model.

Constant delay based on DIVA model: Consider the following delayed neural networks:

$$\begin{cases} x_i(t) = -d_i x_i(t) + \sum_{i=1}^{n} a_{ij} f_j(x_j(t)) + \sum_{j=1}^{n} b_{ij} g_j(x_j(t-\tau_j)) + I_i, \\ x_i(t) = \phi_i(t), t \in [-\tau, 0] \end{cases} \tag{1}$$

Thereinto:

$$\tau = \max_{1 \le j \le n} \{\tau_j\}, j = 1, 2, \ldots, n, n \ge 2$$

represents for the number of neurons, x_i (t) for the statement of ith neuron, d_i>0 for the rate of recovery of isolated resting state with no neurons network or any additional voltage, α_{ij}, b_{ij} for the impact strength from output of jth neuron to ith neuron, f_i, g_i for the output of jth neuron, I_i for the constant external output of ith neuron.

If the following conditions are met to the output functions:

f_j (0) = g_j (0) = 0 and there exists constants α_j, β_j can make any x, y, ∈R, 1 ≤i, j≤n, f_j (x)-f_j (y)|≤α_j |x-y|, |g_j (x) -g_j (y)|≤β_j |x-y|.

In n-dimensional vector space R^n, we define bound norm as following:

$$\|x\|^{\tau} = \sup_{0 \le \delta \le t} \left| \sum_{i=1}^{n} |x_i(s)|^{\tau} \right|, \tau \ge 1$$

Let x^* be an equilibrium point of the system, if there are two constants λ>0, M≥1, it brings:

$$\| x - x^* \| \le M e^{-\lambda t} \| \phi - x^* \|, t \ge 0$$

then x^* is said to be globally stable equilibrium point.

It has proved that the system has only one equilibrium point in a constant delayed network, so the system (1) has only one equilibrium point and the point meets to the global exponential stability conditions.

EXPERIMENT

Model for experiment: The DIVA model has three command modes: Motor Command Mode, FeedForward

Table 1: Attributes for each training object

Learning object	Learning circles	Time	F0	F1	F2	F3
/a/	10	350 m	0~200	650~800	1500~1700	2500~3500
/u/	10	350 m	0~200	200~400	900~1100	2000~2500
/aba/	10	600 m	0~200	650~800	1500~1700	2500~3500
				0~800	0~1700	0~3500
				650~800	1500~1700	2500~3500

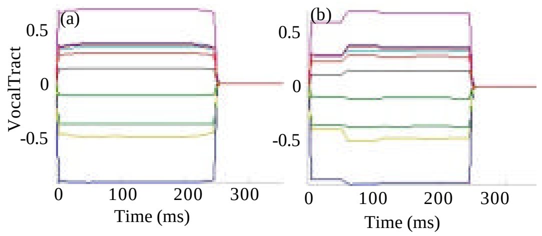

Fig. 2(a-b): Comparison of results of /a/, the left figure shows the experimental results under the original delay mechanism, the right the results under the constant delay mechanism, following figures with arranged

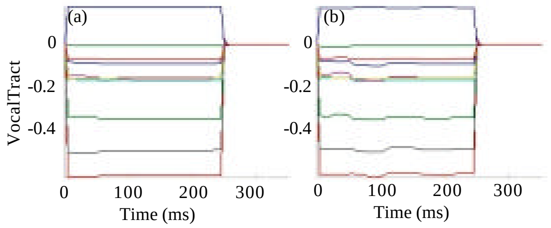

Fig. 3: Comparison of results of /u/

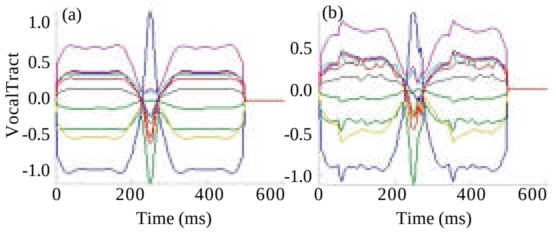

Fig. 4: Comparison of results of /aba/

Command Mode and FeedBack Command Mode. Therefore, we must select some phonemes and syllables for simulation, then test and verify the improvement of application of constant delay mechanism.

In the process of simulation, /a/, /u/ and /aba/ are selected as the learning objects. Some attributes for each training object is shown in Table 1.

Experimental result

Motor command mode: Simulation results are shown in Fig. 2-4.

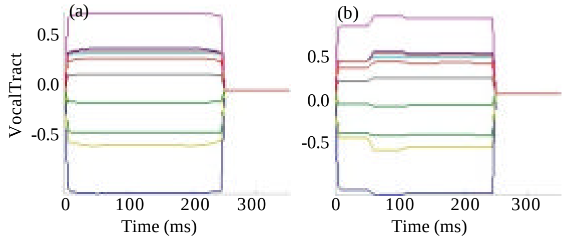

Fig. 5: Comparison of results of /a/

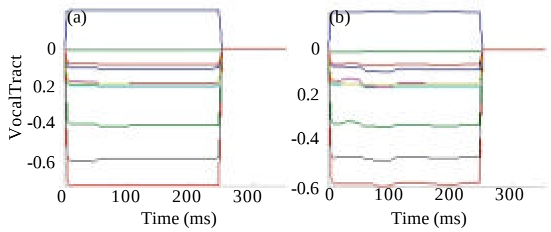

Fig. 6: Comparison of results of /u/

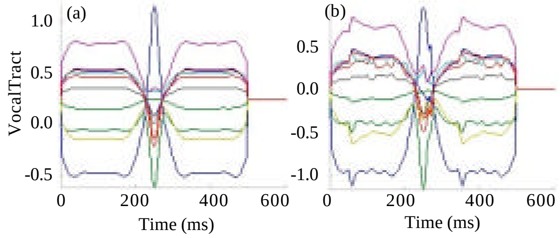

Fig. 7: Comparison of results of /aba/

In motor command mode, trend of vocal process under constant delay mechanism is roughly the same as the trend under original delay model, but closer to the physiological production process and in the experiment we also found the system under constant delay mechanism requires less learning circles to achieve stability, the shortage of it is the sounding curve will have some burr, especially in the polysyllabic case.

FeedForword command mode: The results are shown in Fig. 5-7.

In the FeedForward Command Mode, the simulation results are almost the same as the Motor Command Mode, but we can find the burr was slightly reduced.

FeedBack command mode: The results are shown in Fig. 8-10.

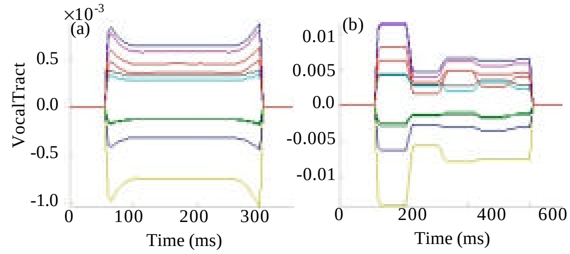

Fig. 8: Comparison of results of /a/

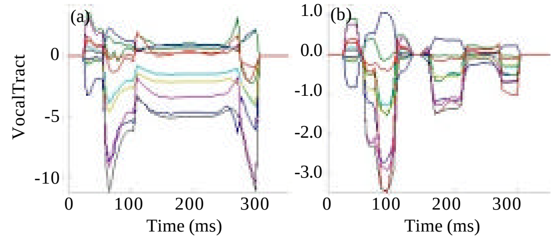

Fig. 9: Comparison of results of /u/

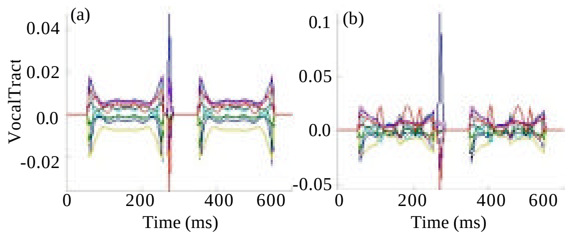

Fig. 10: Comparison of results of /aba/

In the FeedBack command mode, not only the sounding curves are different compared to other modes, but also they change greatly after introducing the constant delay mechanism. it shows that its task complexity greatly rise in the FeedBack command mode, on the other hand it also reflects the differences between the constant delay mechanism and the original, we must do some attributes modification after the application of constant delay mechanism (Qi and Peng, 2010).

CONCLUSION

This study introduced the application of new solutions to the original DIVA model and designs a new system model. What can be different from the existing model is the application of the constant delay mechanism, making the system more consistent with physiological phonation significance in system simulation process sound. But the stability of the system decreases slightly and there are strict conditions for the application of new delay mechanism. The next step will be to further improvement of the existing constant delay mechanism, so it can be improved in terms of stability and adaptability.

ACKNOWLEDGMENTS

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

REFERENCES

Beal, D.S., 2011. The advancement of neuroimaging research investigating developmental stuttering. Perspect. Fluency Fluency Disord., 21: 88-95.

Brown, S., E. Ngan and M. Liotti, 2008. A larynx area in the human motor cortex. Cerebral Cortex, 18: 837-845.

Brumberg, J.S., A. Nieto-Castanon, P.R. Kennedy and F.H. Guenther, 2010. Brain-computer interfaces for speech communication. Speech Commun., 52: 367-379.

Chua, L.O. and L. Yang, 1988. Cellular neural networks: Theory. IEEE Trans. Circuits Syst., 35: 1257-1272.

Golfinopoulos, E., J.A. Tourville, J.W. Bohland, S.S. Ghosh, A. Nieto-Castanon and F.H. Guenther, 2011. fMRI investigation of unexpected somatosensory feedback perturbation during speech. NeuroImage, 55: 1324-1338.

Guenther, F.H., 2008. Neuroimaging of Normal Speech Production. In: Neuroimaging in Communication Sciences and Disorders, Ingham, R.J. (Ed.). Plural Publishing Inc., San Diego, CA., USA., ISBN-13: 9781597561020, pp: 53-85.

Guenther, F.H., S.S. Ghosh and J.A. Tourville, 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang., 96: 280-301.

Qi, Z.Q. and D.L. Peng, 2010. Brain mechanism research of speech processing: Current situation, puzzles and prospect. J. Beijing Normal Univ. (Social Sci.), 4: 40-47.

Tourville, J.T. and F.H. Guenther, 2011. The DIVA model: A neural theory of speech acquisition and production. Lang. Cognit. Processes, 26: 952-981.

Zhang, Y.Y., 2012. Global exponential stability of several cellular neuron networks with delays. Master Thesis, Tianjin University, Nankai, Tianjin, China.