Warlaumont, A. S. (2012). A spiking neural network model of canonical babbling development.
Proceedings of the 2012 IEEE International Conference on Development and Learning (ICDL).
doi: 10.1109/DevLrn.2012.6400842.

A spiking neural network model of canonical babbling development

Anne S. Warlaumont
Cognitive and Information Sciences
University of California, Merced
Merced, CA 95340
Email: awarlaumont2@ucmerced.edu
Web: annewarlaumont.org

Please note the correction at the bottom of p. 6.

***Abstract*—Canonical babbling, the production of vocalizations that contain mature-sounding syllables, is one of the most striking and important milestones prior to the production of first words. This study simulates the emergence of canonical babbling using a spiking neural network containing motor neurons that activate muscles in a vocal tract simulator. The spiking neural network periodically produces synthesized vocalizations and a human listener judges the vocalizations on the basis of their syllabicity, deciding whether or not to reward the model. If a reward is given, spike timing dependent plasticity is increased and the model becomes more likely to recreate a pattern of neural firings similar to that which generated the reinforced vocalization. The model successfully increases its production of mature-sounding canonical syllables, whereas a yoked control simulation does not exhibit any such effect. This finding corresponds to results of experimental work with human infants in which consonant-vowel syllable production is selectively reinforced by the infants' caregivers.**

I. Introduction

A. Canonical babbling in human infancy

One of the most important milestones of vocalization development during the first year is the onset of canonical babbling. Canonical babble refers to a vocalization that contains one or more canonical syllables, i.e. syllables that have acoustic characteristics similar to adult speech [1], [2]. They typically emerge in an infant's repertoire at about 7 months of age. They are preceded by gooing and by marginal babbling, in which vocalizations have the precursors of syllabic structure but the syllables are not well-formed, tending for example to have slower, sloppier consonant-vowel transitions. Development of canonical babbling has been observed to begin with single syllables, e.g. "ba" or "ma," and progress toward sequences of reduplicated syllable types, e.g. "bababa", followed by sequences of variegated syllable types, e.g "bama." The onset of canonical babbling is particularly important because it marks the point at which adult-like phonetic features become relevant and because it provides an essential vocal motor foundation for word production, which begins at about the first birthday. The onset of canonical babbling has been shown to be quite salient to infants' caregivers [3].

Various observational and experimental studies of human infants have pointed toward some of the mechanisms by which canonical babbling develops. Canonical babbling is a very robust phenomenon, affected little by differences in socioeconomic status, language environment, or even mild to moderate hearing impairment [1]. It is, however, delayed by severe deafness [4] and by tracheostomy [5], indicating that both hearing and experience with vocalization are involved in its development. Naturalistic observation has revealed that parents selectively reinforce vocalizations that have mature characteristics, including canonical syllable structure [6], and experimental studies have shown that infants do indeed increase the frequency with which they produce well-timed syllables when they are reinforced specifically for those well-timed syllables [7]. Although not yet studied in human infants (perhaps because it is harder to measure), intrinsic reinforcement, for example where stimuli (including auditory, tactile, and proprioceptive) of moderate complexity are preferred [8], may also play a role in the development of canonical babbling.

B. Existing computational models of infant vocal development

A number of models have been built to account for the development of specific segmental phonetic units, such as particular vowels and consonants. The majority of these models have attempted to model the development of vowel imitation, especially focusing on the role of caregiver input and auditory and perceptual feedback from one's own productions. Many, though not all, have featured non-spiking connectionist neural networks. For example, it has been shown that listening to one's own vocalizations provides useful information for learning to imitate vowels produced by other speakers [9], [10]. When coupled with vowel inputs from other speakers, experience with auditory feedback from one's own random vocal exploration can also facilitate perception of vowels in one's native language [11] and result in a tendency to produce vowels similar to those produced by other proximal individuals [12].

It has also been demonstrated that selectively imitating a model that produces vowels at random when it produces vowels from a target language enables that model to successfully imitate adult vowels, particularly when the infant's own vocal tract is modeled as having a different shape (and thus producing different acoustics) compared to an adult's [13], [14]. Thus, contingent responding by an adult has been shown to be effective in computational models of vowel development, mirroring those findings from the human infant literature that have pointed to an important role of contingent adult responding.

Another model, the DIVA model, has addressed the acquisition of motor control of sequences of both consonants and vowels [15]. The model learns through the experience of receiving sensory (in this case both auditory and tactile) feedback when producing random sequences of vocal tract movements, including movements of articulators such as the lips, jaw, and tongue. It is assumed that the model knows what auditory stimuli correspond to particular phonemes in the model's target language, so that when it happens to randomly produce one of those phonemes, it can form a mapping between the phoneme targets and particular articulatory configurations. After this random exploration, it is given specific phonetic targets and further fine tunes its synaptic connections, learning to produce new speech sounds and sequences of speech sounds. Relatedly, learning to produce particular sequences of vowel sounds based on experienced input sequences has been addressed in work by Kanda et al. [16].

Taken together, these models demonstrate great progress over the past decade in computational modeling of the development of speech production capabilities. However, it appears that no model to date has explicitly addressed the emergence of canonical babbling, that is of the shift from producing vocalizations that lack syllabic structure to producing sequences of vocalizations that do have syllabic structure, in and of itself. The DIVA model comes the closest, having addressed the development of production of consonant-vowel sequences. However, the development of syllabicity independent of the issue of being able to generate specific phonemes given a precise input sequences, has not been explored with that model. Note that human infants babble syllabically many months before they begin to produce sequences of specific phonemes as part of learned word targets as modeled in DIVA [1].

Additionally, all of the models reviewed above utilize either non-spiking neural networks or else even more abstract computational mechanisms. It is useful to model cognitive phenomena such as vocal motor learning at a range of levels of physiological detail vs. abstraction. On the more detailed side, addressing how spiking neurons might generate vocal motor behavior and plasticity would improve the breadth of our understanding of the neural mechanisms underlying speech-related vocal motor development.

C. Spiking neural networks and motor control

Spiking neural networks can be expected to be well suited for modeling infant vocal motor development, particularly modeling the development of canonical babbling. They are naturally temporally oriented, being that the neurons integrate input current over time, spiking (generating action potentials) when a certain voltage threshold is exceeded. Taken together, groups of spiking neurons' voltages or spikes can be summed to generate reasonably realistic traces that have regimes similar to those observed in cortical EEG traces and in vitro recordings [17], [18], [19]. Thus, this class of neural networks provides a natural way to generate temporally rich behavior. These dynamic characteristics are expected to be especially useful since canonical babbling is inherently a temporal phenomenon.

Some models have already been developed that make use of spike timing dependent plasticity (STDP) in spiking neural networks for purposes related to motor control. For example, Bouganis et al. [20] used a spiking neural network with STDP to map spatial targets and proprioceptive information into motor commands for a robotic arm. Their network was strictly feedforward and relied on firing rates to encode both inputs and outputs. Their model demonstrated that a spiking neural network, which is more neurophysiologically realistic than the non-spiking neural network models that have been used in the past for the same arm control task, can also perform well.

In other work, Izhikevich [21] has shown that his spiking neural network model, when equipped with spike timing dependent plasticity that is modulated by the concentration of dopamine in the network, can model several associative and operant conditioning phenomena that are well documented in animal behavior. Of particular interest to the present study, the model can learn to, when given a pulse of input current, respond by activating a certain group of neurons, group A, to a greater extent than another group, group B. The model is simply rewarded when it happens by chance to follow the input pulse with greater group A activation. If after learning to preferentially respond by higher group A activation the contingency of reward is reversed, the model reverses its response. Izhikevich proposed that groups A and B could potentially represent competing muscle groups, in which case the learned response would be a specific motor behavior.

D. Overview

The present study introduces a model of the development of canonical babbling and reduplicated canonical babbling that combines a spiking neural network, a vocalization synthesizer that simulates the human vocal tract, and reinforcement from a human listener. Figure 1 provides a schematic overview of the components of the model.

II. Method

A. The neural network

The neural network used here is Izhikevich's [21] spiking neural network, and was created by modifying the code published on that author's website. Further details can be found there and in [21]. It consists of a network of neurons that received both current from other neurons and a small amount of random external current as input. The neurons spike when a certain threshold membrane potential is reached and exhibit a refractory period after spiking. 800 excitatory and 200 inhibitory neurons are included. Each neuron in the model is connected to a randomly selected set of 100 other neurons with, for simplicity, a transmission delay of 1 ms (although in the future variable delays could be tried).

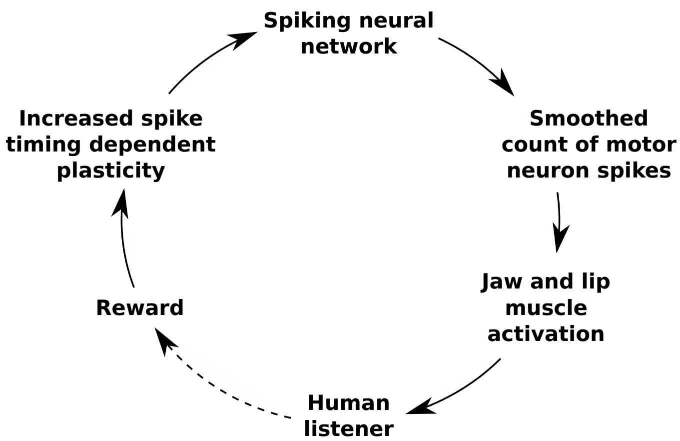

Fig. 1. Schematic overview of the steps involved each time the model vocalizes. The loop begins with simulation of the spiking neural network, pictured at the top. The dashed line indicates that sometimes the human decides not to reward the model for its vocalization, in which case there is no increase in spike timing dependent plasticity.

The neural network learns via spike timing dependent plasticity (STDP), which works as follows. When one neuron (neuron A) fires followed by firing of a second neuron (neuron B), the connection from neuron A to neuron B (if there is one) is strengthened and the connection from neuron B to neuron A (if there is one) is weakened. The amount of strengthening or weakening of synapses decreases exponentially with the time between the firing of the two neurons.

The STDP is dependent on the presence of dopamine, with there always being a constant low level of dopamine and there being surges of greater amounts of dopamine when the neural network receives reinforcement for its behavior. There is an decaying STDP eligibility trace such that when a dopamine surge occurs very quickly after a particular pattern of neuronal firing, a relatively large change is made to the synaptic weights and when the dopamine surge occurs quite awhile after the pattern of firing, less learning takes place.

The neural network runs in 1 ms (simulated time) steps, and was run in this study for a total of 2 hours of simulated time.

B. From spikes to muscle activations

Fifty of the excitatory neurons in the network were randomly chosen to be motor neurons. Every five seconds of simulation time, the spikes of these motor neurons over the course of the subsequent second were used in order to generate jaw and lip muscle activations, which were in turn used to generate a vocalization. In other words, one fifth of the time the model's behavior was translated into a vocalization, listened to, and either reinforced or ignored. The other four fifths of the time the model's behavior was ignored. At each 1 ms time step within the 1s vocalization period, the number of motor neurons that fired was counted (the max possible at any given time step was fifty). Then a 100 ms moving average of these spike counts was obtained, resulting in a 900 ms series of smoothed motor neuron spike counts. See Figure 2A to see examples of these moving averages.

C. Vocalization synthesis

A vocalization was then synthesized using Boersma's articulatory synthesizer [22]. The synthesizer models the human vocal tract as a series of air-filled tubes, where the walls of the tubes are made up of many coupled mass-spring systems. It was chosen because it is conveniently available through Praat [23] and because it is well-suited to modeling consonant production, whereas the most commonly used realistic vocal tract synthesizer, Maeda's [24], must be specially modified in order to produce consonants [15]. Here the adult female version of the vocal tract was used. The user specifies the duration of the vocalization and the lung volume and muscle activations at various times within that duration. The synthesizer linearly interprets lung volume and muscle activations between these user-set points.

For all vocalizations in the present study, the duration was 900 ms, corresponding to the length of the smoothed motor neuron spike count time series. Lung volume was set to be .1 from 0 to .02 ms and then to reduce to zero starting at .05 ms until the end of the vocalization. Additionally, the interarytenoid, a laryngeal muscle that promotes phonation (the generation of sound in the larynx) was set to a constant value of .3. The hyoglossus, a muscle that pulls the tongue down, was also set to a constant value of .3, in order to promote the production of the vowel /a/. These lung, interarytenoid, and hyoglossus parameters were set to be similar to those used in Boersma [22] to model the production of the sound "aba".

In that "aba" example, the masseter jaw muscle and the orbicularis oris lip muscle were also manipulated. Activation of the masseter muscle affects the equilibrium position of the jaw, promoting jaw closure. Activation of the orbicularis oris muscle affects the equilibrium position of the lips, promoting lip closure. Here too we manipulate these muscles but rather than using the activations provided in Boersma's example, the masseter and orbicularis oris muscles were set dynamically based on the outputs of the motor neurons in the spiking neural network: Each 1 ms element of the 900 ms smoothed spike count time series was multiplied by 10 then sent to the synthesizer as both masseter and orbicularis oris activations.

Using fluid dynamics and the mechanics of the mass-spring walls of the vocal tract, the synthesizer calculates the pressure of the air within the simulated vocal tract. The time series of pressure at the lips gives the sound that is output by the synthesizer.

D. Human listener reinforcement

As was mentioned above, a 900 ms vocalization was produced every 5 s. After the motor neuron spikes used as muscle activations were generated, the vocalization was synthesized and then the simulation was paused to allow the human listener (the author) to play the vocalization sound. After playing the sound, the listener was prompted to enter "1" to provide the network with a reward or to enter "0" to withhold reward. The listener's goal was to get the model to learn to produce canonical syllables, that is, syllables with distinct consonant-vowel or vowel-consonant transitions, by responding when

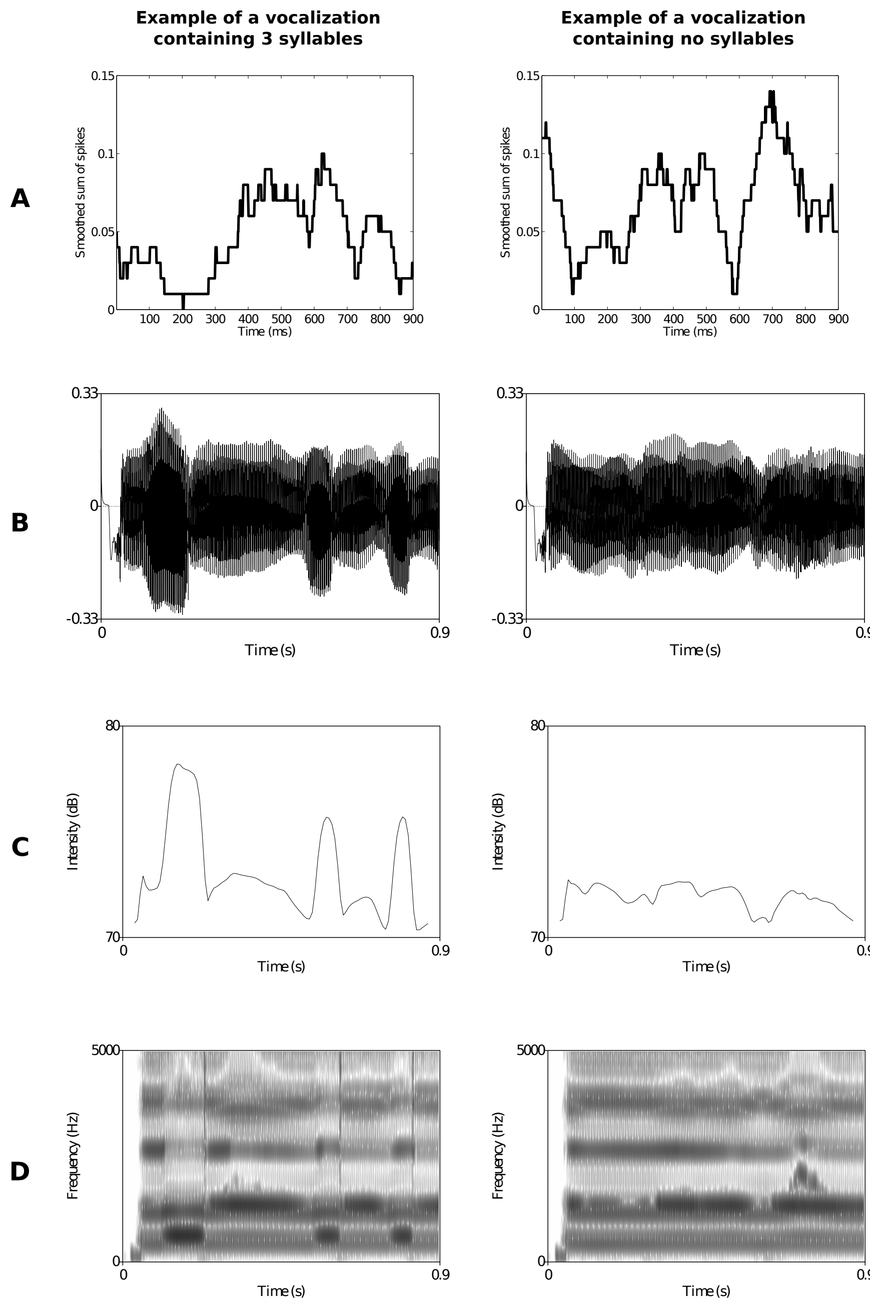

Fig. 2. Examples of vocalizations produced by the model. On the left is an example vocalization that had a high canonical babble quality score, containing three well-formed canonical syllables, and on the right is an example that had a low canonical babble quality score, containing no well-formed canonical syllables. A: Jaw (masseter) and lip (orbicularis oris) muscle activations, given by obtaining a moving average, over 100 simulated milliseconds, of firing counts for a subset of fifty neurons designated to be motor neurons. B: The sound waveforms generated by the vocal tract simulation based synthesizer. C: Amplitudes of the vocalizations. D: Spectrographic plots of the vocalizations.

the model produced vocalizations containing such syllables. Particular preference was given to sequences of more than one canonical syllable. In this way, the human listener acted as a sort of parent to the model, responding when it produced something particularly mature-sounding.

If the human listener provided the model with a reward for the just-produced vocalization, a surge of dopamine was provided to the neural network at the time corresponding to the end of the vocalization. The dopamine surge increased spike timing dependent plasticity, making it more likely that the particular patterns of neural firing involved in that vocalization would recur in the future. In total, the human listener listened to 1,440 vocalizations over the course of the 2 hours of simulated time in the experiment (which actually took several days of real time).

E. Yoked control

Since STDP can potentially affect the neurons' firing patterns even if that STDP is not tied directly to a particular target behavior [18], a control simulation was run. The control simulation was exactly the same as the main simulation except that which neurons were connected was different due to their being randomly assigned, and external current to each neuron at each time step was also different since that too was randomly generated. The control simulation, like the main simulation, produced a vocalization every 5 s. However, its reward and therefore its dopamine release was not based on human listener judgment but was given on a schedule that was yoked to that of the main simulation. Thus, the yoked control simulation was rewarded just as often as the main simulation, but the reward did not bear any systematic relationship to the vocalizations produced by the network or to the patterns of neuronal firings that generated those vocalizations.

III. Results

To evaluate the the model's learning, the human listener listened to every vocalization produced by both the main simulation and the yoked control simulation during the course of their training (2,880 vocalizations in total). The vocalizations were randomly scrambled so that the listener did not know which simulation generated a given sound or when during the course of learning it was produced. Each vocalization was given an integer score between 1 and 4 based on the perceived quality of the vocalization with respect to its syllabicity or lack thereof. A score of 1 tended to be given to sounds that contained no canonical or nearly canonical syllables. A score of 2 tended to be given to sounds that had exactly one canonical or nearly canonical syllable or to sounds that had multiple syllables but where those syllables were low quality. A score of 3 tended to be given to sounds with two canonical or nearly canonical syllables, and a score of 4 tended to be given to sounds with more than two canonical or nearly canonical syllables, or to sounds with two very high quality canonical syllables. Thus, higher scores corresponded to sound that were perceived as containing more, higher quality syllables.

Figure 2 shows two vocalization examples, one that was given a score of 4 and contained three distinct canonical syllables and one that was given a score of 1 and contained a bit of subtle modulation but no canonical syllables. The muscle activations (which are based on smoothed counts of motor neuron firings), the resulting synthesized acoustic waveform, the amplitude envelope of that acoustic waveform, and a spectrogram are shown for each example. It is worth noting that the translation from motor neuron spikes, and therefore from muscle activations, to the amplitude and spectral features of the resulting sound appears to be rather complex and nonlinear. Some changes in muscle activation cause salient syllabic consonant-vowel transitions whereas other, seemingly equally large, changes in muscle activation do not lead to salient syllabicity.

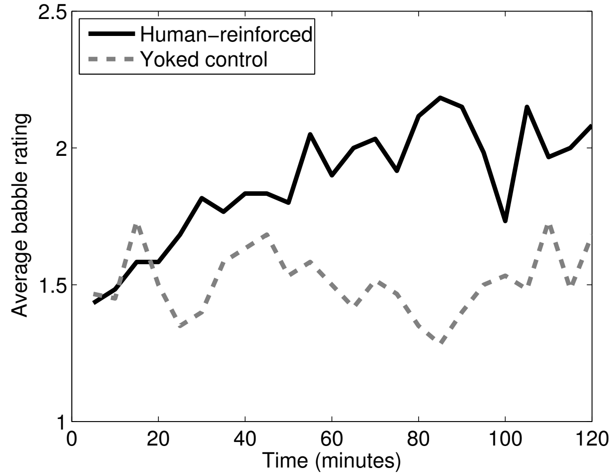

Fig. 3. Canonical babble quality ratings of the model's vocalizations across simulated time as judged by a human listener. The listener was blind to which simulation produced each vocalization as well as to when it was produced.

Figure 3 shows the change in vocalization quality rating as a function of simulation time. The main simulation, in which reward was given by a human listener based on the syllabic structure of the vocalization, produced increasingly high quality babbling over the course of the simulation. The correlation between vocalization quality score and simulation time was highly statistically significant, $r = .19$, $p < .001$. In contrast, the yoked control simulation's vocalizations did not exhibit any significant change in the syllabic quality of its vocalizations over time, $r < .01$, $p = .86$.

IV. Conclusion

This study introduced a new model of the emergence of canonical babbling in infancy. The onset of canonical babbling is recognized to be a major milestone of vocal development in human infancy, providing an essential foundation for the production of first words. This appears to be the first model to address how the tendency to produce vocalizations with syllabic structure develops. It is also almost certainly the first to combine a spiking neural network, a vocal tract simulation based sound synthesizer, and interaction with a real human listener.

Over the course of the two hour simulation, the model developed a greater tendency to produce higher quality canonical babble. In a control simulation, the model's rewards were not contingent on syllabic productions and the model did not exhibit any increase or decrease in production of canonical babble. This result corresponds quite nicely to Goldstein and Schwade's [7] finding that when infants were rewarded socially by their mothers for producing well-timed consonant-vowel structures, they subsequently did produce more of those reinforced sounds. In contrast, when infants' social rewards were yoked to the reward schedule of another infant and were therefore not contingent on any aspect of their own vocal behavior, no change in vocal behavior was observed.

The model's increase in syllabicity quality over the course of learning was fairly gradual, as shown in Figure 3. This result corresponds to the finding that human infants develop canonical babbling seemingly somewhat gradually, moving from non-syllabic vocalizations or very primitive goos to marginal babbling to canonical babbling, and moving from producing predominantly single canonical syllables to producing sequences of reduplicated canonical syllables. More specific judgments regarding the types of syllables produced by the model and of the numbers of syllables in its utterances would be worth tracking separately in the future.

Since the model focuses on addressing a phenomenon that hasn't yet been addressed by other models of infant vocal motor development, it is not possible to directly compare its performance to that of other models. In the future, variations on the model should be tested against each other. Additionally, a model like DIVA that uses non-spiking neural networks might be usefully adapted to see if it can also model the emergence of canonical babbling in the vocal repertoire.

Numerous additional future directions would also be pursuing, given the good performance of the model presented here. For one, the model should be run using several different human listeners for both training and testing. This is currently in progress. Additionally, investigation of neural network weights and of muscle outputs over the course of learning would lend greater insight into how the neural network is adapting. An attempt to include a larger number of muscles (including lung muscles), and to model them using separate groups of neurons, would be more realistic and would make the model more relevant to word learning. Additionally, it is possible that reinforcement could derive not only from social sources but also from intrinsic ones; it would be worthwhile to examine the extent to which this occurs in human infancy as well as explore whether the same neural mechanism could underlie learning from both sources of reinforcement. Likewise, the role of priopriocеption and of hearing oneself as well as others could also be integrated. Finally, if the model, especially the vocalization synthesis component, can be sped up to real-time speed, such a model could conceivably be incorporated into robots that interact more richly with humans.

ACKNOWLEDGMENT

This work was supported by the Department of Energy Computational Science Graduate Fellowship Program of the Office of Science and National Nuclear Security Administration in the Department of Energy under contract DE-FG02-97ER25308.

REFERENCES

[1] D. K. Oller, *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
[2] D. K. Oller and M. P. Lynch, "Infant vocalizations and innovations in infraphonology: toward a broader theory of development and disorders," in *Phonological development: Models, research, implications*, C. A. Ferguson, L. Menn, and C. Stoel-Gammon, Eds. Timonium, MD: York Press, 1992, pp. 509–536.
[3] D. K. Oller, R. Eilers, and D. Basinger, "Intuitive identification of infant vocal sounds by parents," *Developmental Science*, pp. 49–60, 2001.
[4] D. K. Oller and R. E. Eilers, "The role of audition in infant babbling," *Child Development*, vol. 59, pp. 441–449, 1988.
[5] J. L. Locke and D. M. Pearson, "Linguistic significance of babbling: evidence from a tracheostomized infant," *Journal of Child Language*, vol. 17, no. 01, pp. 1–16, 2009.
[6] J. Gros-Louis, M. J. West, M. H. Goldstein, and A. P. King, "Mothers provide differential feedback to infants' prelinguistic sounds," *International Journal of Behavioral Development*, vol. 30, no. 6, pp. 509–516, 2006.
[7] M. H. Goldstein and J. A. Schwade, "Social feedback to infants' babbling facilitates rapid phonological learning," *Psychol. Sci.*, vol. 19, no. 5, pp. 515–523, 2008.
[8] C. Kidd, S. T. Piantadosi, and R. N. Aslin, "The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex," *PLoS ONE*, vol. 7, no. 5, p. e36399, 2012. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0036399
[9] I. Heintz, M. Beckman, E. Fosler-Lussier, and L. Ménard, "Evaluating parameters for mapping adult vowels to imitative babbling," in *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Brighton, UK, 2009.
[10] A. S. Warlaumont, G. Westermann, and D. K. Oller, "Self-production facilitates and adult input interferes in a neural network model of infant vowel imitation," in *AISB 2011 Computational Models of Cognitive Development*, D. Kazakov and G. Tsoulas, Eds. York, UK: Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 2011, pp. 8–12.
[11] G. Westermann and E. R. Miranda, "A new model of sensorimotor coupling in the development of speech," *Brain and Language*, vol. 89, no. 2, pp. 393–400, 2004.
[12] P.-Y. Oudeyer, "The self-organization of speech sounds," *Journal of Theoretical Biology*, vol. 233, no. 3, pp. 435–449, 2005.
[13] Y. Yoshikawa, M. Asada, K. Hosoda, and J. Koga, "A constructivist approach to infants vowel acquisition through mother-infant interaction," *Connection Science*, vol. 15, no. 4, pp. 245–258, 2003.
[14] K. Miura, Y. Yoshikawa, and M. Asada, "Vowel acquisition based on an auto-mirroring bias with a less imitative caregiver," *Advanced Robotics*, vol. 26, no. 1-2, pp. 23–44, 2012.
[15] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, "Neural modeling and imaging of the cortical interactions underlying syllable production," *Brain Lang.*, vol. 96, no. 3, pp. 280–301, 2006.
[16] H. Kanda, T. Ogata, T. Takahashi, K. Komatani, and H. Okuno, "Continuous vocal imitation with self-organized vowel spaces in recurrent neural network," in *2009 IEEE International Conference on Robotics and Automation*, Kobe, May 2009, pp. 4438–4443.
[17] E. M. Izhikevich, "Simple model of spiking neurons." *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1569–1572, 2003.
[18] ——, "Polychronization: Computation with spikes," *Neural Computation*, vol. 18, no. 2, pp. 245–282, 2006.
[19] A. S. Warlaumont, H. C. Lee, M. Benayoun, R. L. Stevens, and M. Hereld, "Systematic comparison of the behaviors produced by computational models of epileptic neocortex," *Journal of Clinical Neurophysiology*, vol. 27, no. 6, pp. 479–483, 2010.
[20] A. Bouganis and M. Shanahan, "Training a spiking neural network to control a 4-DoF robotic arm based on spike timing-dependent plasticity," in *Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2010, pp. 1–8.
[21] E. M. Izhikevich, "Solving the distal reward problem through linkage of STDP and dopamine signaling," *Cerebral Cortex*, vol. 17, no. 10, pp. 2443–2452, 2006.
[22] P. Boersma, *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics, 1998. [Online]. Available: http://www.fon.hum.uva.nl/paul/papers/funphon.pdf
[23] P. Boersma and D. Wennink, "Praat: Doing phonetics by computer (Version 5.1.31) [Software]. Available from http://www.praat.org," Apr. 2010. [Online]. Available: http://www.praat.org
[24] S. Maeda, "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech production and speech modeling*, W. J. Hardcastle and A. Marchal, Eds. Boston: Kluwer Academic Publishers, 1990, pp. 131–149.

Correction: After publication, it was discovered that the model did not receive reinforcement in the way that the paper described. The main conclusions still hold after correcting the error and making the following changes: performing synaptic weight normalization in place of spike-timing-dependent depression, removing the constant low level of dopamine, and scaling muscle activity by 2.5. A simulation and yoked control of the corrected model exhibited an increase in performance over time in the main simulation ($r=.096$, $p<.001$) as well as, to a lesser extent, in the control ($r=.070$, $p=.008$), and performance was significantly higher overall in the main simulation compared to control ($T=1.870$, $p=.031$).