Modeling Motor Pattern Generation in the Development of Infant Speech Production

Ian Spencer Howard[1] & Piers Messum[2]

[1]Computational & Biological Learning Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK
[2]Centre for Human Communication, University College London, London WC2E 6BT, UK
E-mail: ish22@cam.ac.uk, p.messum@ucl.ac.uk

Abstract

We previously proposed a non-imitative account of learning to pronounce, implemented computationally using discovery and mirrored interaction with a caregiver. Our model used an infant vocal tract synthesizer and its articulators were driven by a simple motor system. During an initial phase, motor patterns develop that represent potentially useful speech sounds. To increase the realism of this model we now include some of the constraints imposed by speech breathing. We also implement a more sophisticated motor system. Firstly, this can independently control articulator movement over different timescales, which is necessary to effectively control respiration as well as prosody. Secondly, we implement a two-tier hierarchical representation of motor patterns so that more complex patterns can be built up from simpler sub-units. We show that our model can learn different onset times and durations for articulator movements and synchronize its respiratory cycle with utterance production. Finally we show that the model can pronounce utterances composed of sequences of speech sounds.

1 Introduction

Our previous work modeled a non-imitative account of the development of infant speech production [1-3]. The model runs firstly as a stand alone system and then interacts naturally with a caregiver. We showed that rewarded exploration of the vocal tract leads to the discovery of potentially useful vocal motor schemes [4] (potential speech sounds). Imitative exchanges between the infant and caregiver, involving reformulations of the infant's output, lead to the association of the infant's motor actions to the adult judgment of their linguistic value expressed in a vocal form. This solves the correspondence problem for the sub-word units that are used in learning the pronunciation of words (with the exception of the very first words). This enables the infant to learn words by imitation, and it is then taught the names of objects by the caregiver.

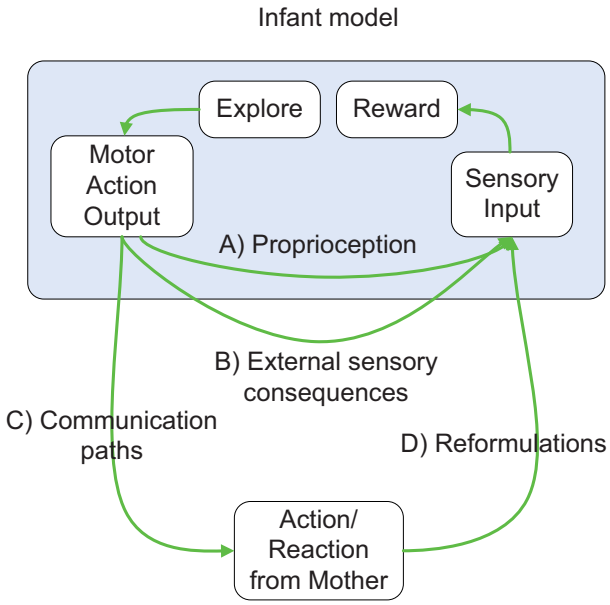

Figure 1: Agent model of an infant showing the important signal flow paths with a female caregiver.

This earlier work concentrated on the most important principles in our account, such as the

reinforcement and reformulation signal flow paths between the infant's motor output and his sensory inputs (figure 1). It employed a Maeda vocal tract synthesizer [5] and used simple mechanisms of perception, production and association. In the motor generator a basic motor pattern was defined in terms of starting and ending articulator target positions. The dynamics of the vocal tract system determined the path of the movement between the targets, and the trajectories were computed by interpolating between sequential targets using critical damping [6]. Although sufficient for the purposes of our initial demonstration, the motor generator suffered from some limitations which we address in this paper. We now also begin to model speech breathing.

2 Modeling speech breathing

Previously we ignored the role of the respiratory system in speech production but we have now taken the first steps towards accounting for its effects. Voluntary control of speech breathing is implemented by specifying a lung compression force to determine the direction and rate of airflow. At any given time, lung volume and the limits of normal lung volume excursion constrain how much air is available to support sound production.

We take the vital capacity of the infant's lungs to be about 25% of adult values. In adults, exhalation is largely passive, driven by the elasticity of the pulmonary/chest wall unit, but this is not the case in infants [7], so we do not include a spring term in our model. When lung volume reaches the upper or lower limits, airflow is set to zero. Voicing is only permitted during the expiratory phase. Breathing thus affects the utterances that can be generated and speech production will be disrupted if it is not synchronized with articulator movements. It is therefore one task of the learning mechanism to synchronize breathing with the movement of the articulators so that useful utterances can be reliably produced.

3 Articulator movements

In our previous work all the articulators moved synchronously between defined targets. In reality speakers are able to control different articulator movements independently over different timescales, starting points and durations, as demonstrated by the movements used to generate syllables compared with those which create intonation contours, which span multiple syllables.

We now give each articulator its own starting offset time and duration of movement. This ability to independently control aspects of the speech production apparatus is necessary for us to be able to introduce speech breathing into the model.

4 Construction of sequences

A second extension of the motor system involves the ability to learn and produce temporal sequences of sub-sounds. This is an important capability since words (and sentences) are constructed from smaller sub-patterns. The model automatically segments input speech syllables spoken by a male subject provided they are separated by silence (allowing segmentation to be carried out on the basis of acoustic power). By recognizing these sub-sounds and using the corresponding sequence of motor actions, the model could learn to produce multi-syllable words by imitation.

5 Results

We re-ran our original experiments to discover simple potential speech sounds based on salience, and then established a linguistic value for some of these through caregiver reformulations. This time we included constraints due to breathing and we incorporating the improved motor system using motor patterns that independently specify independent timings for each articulator transition.

Figure 2 shows a speech sound found by exploration, together with the associated salience and effort terms that were used to determine the reward for the optimization procedure. Asynchronous articulator movement was apparent in an increase in the variety of basic sounds discovered. The ability to asynchronously move articulators is manifested in their trajectories and the air flow.

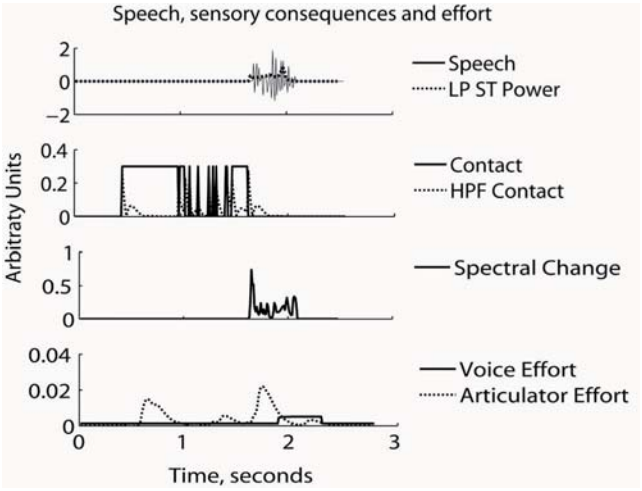

Figure 2. Example of discovered speech sound and its evaluation in terms of salience and reward.

Figure 3 shows an example of a good speech utterance discovered by the model. Notice also the asynchronous movement of the articulators and the breathing trajectories. In this example, expiration coincides with when voicing is useful, resulting in a good speech sound. In the lower plot the air flow gates the preliminary voicing-without-accounting-for-airflow signal (VxWAF) resulting in a reduced duration of voicing control to the synthesizer (Vx). However it does so in a fashion consistent with generating a useful utterance. That is, the model breathes in before a syllable and out during it. There is also a no-flow pause in-between when the lungs have filled to the upper limit of their normal excursion. Good breathing synchronization does not always occur and some sounds generated are not useful speech utterances because of the inappropriate coordination of articulator movement and the air flow, as shown in figure 4. Here the articulations show poor synchronization with breathing.

6 Discussion

We added a model of speech breathing to an articulatory synthesizer. Speech breathing plays an important role in the development of pronunciation [3, 9] but is often omitted from computational models of speech production. In addition we built a more sophisticated motor system which implemented asynchronous articulator control over different timescales on different articulators.

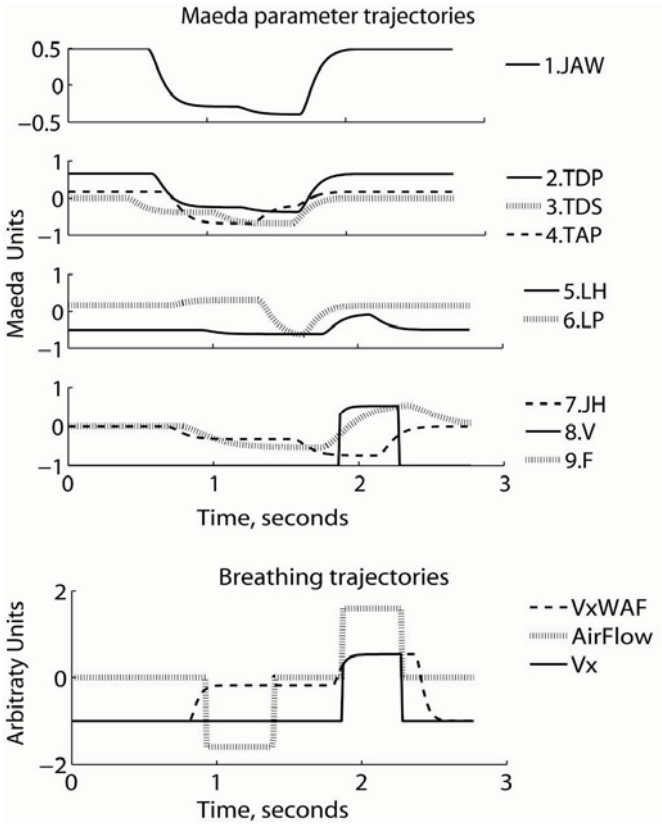

Figure 3. Trajectories for Maeda synthesizer parameters and breathing, for a good utterance

The ability to control movements over independent time scales is important because breathing operates over a much longer time scale than the movements involved in the realization of phonetic contrasts. The value of this was demonstrated in the model's ability to coordinate breathing control with articulator movement. In the future we will look at the issues arising from the coordination of breathing with multi as well as single syllable utterances without our model having to independently relearn all breathing timings for each different case. This will also involve the control mechanism taking notice of the different airflow requirements of different syllables.

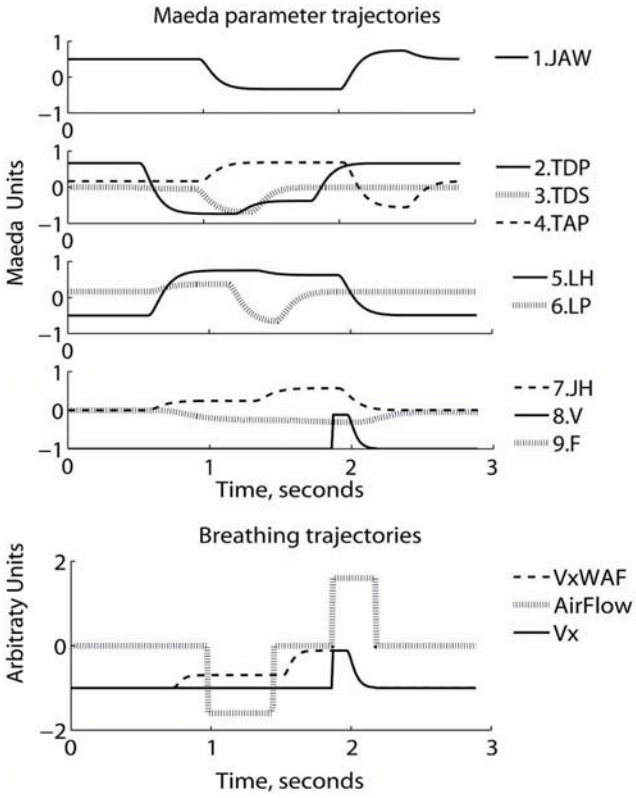

Figure 4. Maeda synthesizer and breathing trajectories for a poor utterance

Furthermore, the more sophisticated hierarchical temporal pattern generator we implemented enables our model to learn multi-syllable words by stringing together single syllables in sequence (see online supplementary information for examples as .WAV files).

Prosody (the timing, stress and intonation of speech) is important in human communication since it conveys some of the speaker's linguistic meaning as well as his emotional state. Prosodic aspects of speech manifest themselves over a longer timescale than phonetic contrasts, so long scale coordination of motor activity is required to control them. There is increasing understanding of the role played by embodiment in the development of complex systems [8]. Taking note of this, if we are to model speech development we should recognize that an infant's vocal apparatus is not simply a scaled down version of an adult's. Speech breathing and speech aerodynamics play an important role in infant speech development [3, 9] and our model's modified motor system can now begin to deal with these issues.

Additional information including examples of the model's speech output is available online at: www.ianhoward.info/issp_2008.htm

7 References

[1] Howard, I.S. and P.R. Messum, *Modeling infant speech acquisition using action reinforcement and association.* In *Speech and Computer, SPECOM 2007.* Moscow Linguistics University, 2007.

[2] Howard, I.S. and P. Messum, *A computer model that learns to pronounce using caregiver interactions.* Forthcoming.

[3] Messum, P.R., *The Role of Imitation in Learning to Pronounce*, PhD thesis, London University, 2007.

[4] McCune, L. and M.M. Vihman, *Vocal Motor Schemes*. In *Papers and Reports in Child Language Development,* Stanford University Department of Linguistics 26, 72-79. 1987.

[5] Maeda, S., *Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model.* In *Speech production and speech modelling.* W.J. Hardcastle and A. Marchal, Editors. Kluwer Academic Publishers: Boston, 1990.

[6] Markey, K.L., *The sensorimotor foundation of phonology; A computational model of early childhood articulatory development.* PhD thesis, University of Colorado: Boulder-Colorado, 1994.

[7] Netsell, R., et al., *Developmental patterns of laryngeal and respiratory function for speech production.* J Voice, 8(2), 1994.

[8] Sirois, S., et al., *Precis of neuroconstructivism: how the brain constructs cognition.* Behav Brain Sci 31(3), 2008.

[9] Messum, P., *Embodiment, not imitation, leads to the replication of timing phenomena.* In *Proceedings of Acoustics '08*, 2405-2410 Paris: ASA, 2008.