A Computational Model of Infant Speech Development

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

[1]Biological & Machine Learning Lab, Department of Engineering, University of Cambridge, Trumpington Street,
Cambridge CB2 1PZ, England,
[2]Department of Phonetics & Linguistics, University College London, Gower Street, London WC1E 6BT, England
E-mail: ish22@cam.ac.uk, p.messum@ucl.ac.uk

Abstract

Almost all theories of child speech development assume that an infant learns speech sounds by direct imitation, performing an acoustic matching of adult output to his own speech. Some theories also postulate an innate link between perception and production. We present a computer model which has no requirement for acoustic matching on the part of the infant and which treats speech production and perception as separate processes with no innate link. Instead we propose that the infant initially explores his speech apparatus and reinforces his own actions on the basis of sensory salience, developing vocal motor schemes [1]. As the infant's production develops, he will start to generate utterances which are sufficiently speech-like to provoke a linguistic response from its mother. Such, interactions are particularly important, because she is better qualified than he is to judge the quality of his speech. Her response to his vocal output is beneficial in a number of ways. Because she is a learned speaker, her experienced perceptive system can effectively evaluate the infant's output within the phonological system of the ambient language L1. Simply generating a salient response will tend to encourage the infant's production of a given utterance. More significantly, during imitative exchanges in which the mother reformulates the infant's speech, the infant can learn equivalence relations using simple associative mechanisms between his motor activity and his mother's acoustic output, and thus can solve the correspondence problem. Notice that the infant does not learn equivalence relations between his own acoustic output and that of his mother based on acoustic similarity. Any similarity based matching need only needs to be performed by his mother.

We present the results from preliminary experiments and demonstrate that this model is able to progress through two distinct stages of speech development. It begins by generating simple sounds and ends up producing word-like utterances.

1. Introduction

There are currently several general models of how an infant learns to speak e.g. [2] including some computational ones [3,4,5]. Most of them assume that he learns to produce speech sounds by imitating those heard in the linguistic environment either using an innate mapping between perception and production [6], or using an acoustic match between what he hears and what they produce, such as some previous work [7,8]. These models are problematic for a number of reasons [9] and many of the computational ones do not address the normalization problem between infant and adult speech. There is no evidence that infants learn the qualities of speech

sounds by direct imitation, although words and phrases may be mimicked as whole shapes in the early stages of acquisition [9]. We argue that imitation only becomes useful as a mechanism after the basic level of perceptual and productive skills have first been acquired by other means. This view is supported by the observation that if learning was performed by acoustic matching, one would expect a continuous improvement of speech sound quality over time, and this is not the case. Rather, speech sounds are often only modified when they are not accepted by the infant's mother and consequently do not evoke a desirable response. Our model takes the approach that phonological development emerges from an interaction between auditory perception and motor control, following [10, 11]. We treat speech production and perception as independent systems and formulate speech production development as an optimization problem which involves finding good motor patterns that can drive the vocal apparatus in such a way as to maximize reward. To do so, we argue that an infant's perception and production need only initially interact with each other through reinforcement pathways, and that perception does not need to provide a detailed production target for the learning of speech sounds to occur. As speech production ability develops, the infant will, however **learn to imitate**, because he will learn associations between phonologically significant sounds and his realization of them. This quickly leads to a more efficient way of learning word forms than mimicry of whole shapes: words can be broken down into sequences of speech sound elements within the infant's repertoire, which then activate his associated motor patterns and thus generate acoustic output.

2. The Computational Model

2.1. Learning speech production and their associations

To be able to speak, we must have learned abstract patterns of neural motor activity that represent corresponding sequences of movements of the vocal apparatus. We must also have learned the motor skills needed to move our articulators with sufficient precision to generate speech sounds accurately and reliable, and these two levels of control are intimately linked and develop together. Much of the vocal apparatus is used for respiration, feeding and crying, etc, so some actions needed to generate speech output will develop from birth. Such autonomous patterns of behavior will need be refined further and controlled voluntarily to support speech production and we may initially expect interaction artifacts to occur until the necessary speech skills have been mastered. The neural mechanism needed to control the vocal apparatus will need to deal with noise in the motor neurons, variable physical properties of the muscles and vocal apparatus as well as any

physical disturbances and perturbations, and to do so can take advantage of all forms of proprioceptive, tactile and auditory feedback. The motivation for the infant in our model to learn to speak is to generate reward, either directly due to the sensory consequences of its own actions (see figures 1,2) or indirectly due to a response from his mother (see figures 3,4). If we build the reward system to be more sensitive to externally generated stimuli (which will occur if we attenuate the sensory consequences of its own actions [12,13]) the mother's response will be especially highly rewarded. Using this framework, the spontaneous generation of speech utterances, initially independent of any internal state of mind of the infant, will get rewarded if they generate sensory consequences. As utterances begin to sound more interesting and maybe speech like to the mother, we assume that her probability of generating a response increases.

2.2. Reinforcement learning optimization framework

Representing learning speech production as an optimization problem, let us describe the abstract motor patterns in terms of a set of parameters P, the mapping between motor pattern to articulator positions as M, and the reward for the resulting speech sound as R. We note that mapping M is effectively a controller that maps an abstract representation of motor movements via low level muscle activations to positions and is essentially independent of any given pattern. It should therefore be optimized over all sets of movement patterns. In addition, this controller should operate closed-loop and take advantage of proprioceptive as well as tactile and acoustic feedback to maintain robust operation in the presence of disturbances of the vocal apparatus. Within this framework, learning a speech sound corresponds to finding pattern parameters P and optimizing the mapping M, that maximize reward R. Because reward is an evaluation of performance, and does not represent a target to be achieved, we note that this scheme operates in a reinforcement learning framework, as opposed to a supervised one. Several questions naturally arise from this formulation. What are the parameters P, what is the mapping M, how do we define the reward R and how do we use it to find the optimal parameters P that represents a desirable speech utterance? If the production system makes exploratory movements, it will be possible to discover useful motor patterns by reinforcing those that result in speech-like sounds, and this will lead to the discovery of what McCune and Vihman describe as useful vocal motor schemes, which are generalized action patterns that yield consistent phonetic forms [1].

All languages make use of sounds created by the basic gestures which emerge as a consequence of making simple movements of the vocal apparatus and this point also highlights the importance of using an accurate model of the vocal apparatus if we wish to explain the developmental stages of infant speech production. The physical structure (its kinematic and dynamic properties, but also its aerodynamic ones [9]) places constraints on the position and trajectories the articulators can take. This affects what kinds of static and dynamic sounds can be generated by simple movements, and also how easily. It is also important to notice that the simple idea of performing optimization of the motor patterns M based on reward will lead to learning, and many observations of the speech acquisition process will be epiphenomenal and emerge

as a consequence of the motor system learning to control this physical system, rather than representing underlying goals themselves.

2.3. Signal flow paths and reward

One key issue to our discussion here is the source of the reward R and this relates to the sources of feedback available to the infant from which the quality of his speech production can be evaluated

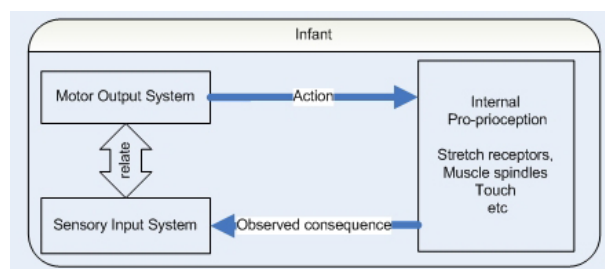### 2.4. Direct feedback of sensory consequences

Figure 1: Signal flow interactions showing the direct body-internal sensory consequences of his action.

When the infant makes a speech sound, he will receive internal proprioceptive feedback due to articulator movement and contact, as well as from the muscle spindles. There will be internal feedback within his head due to vibration from vocal folds, the turbulence of the air in the vocal tract, and this signal flow pathway is shown in figure 1. The infant will hear the acoustic disturbance that arises using his hearing apparatus, although there is evidence that acoustic feedback is heavily degraded by bone conduction, making it of limited value for assessing speech production. He will also be able to hear any effect his speech is having via his external environment (figure 2). The amount of motor activity will also result in a cost for the action, which will reduce reward. Even infants with a naive perceptive system will be able to judge the salience of this proprioceptive and acoustic feedback. During initial development of our model, we will used these signal flow paths as the dominant source of reinforcement to speech production.
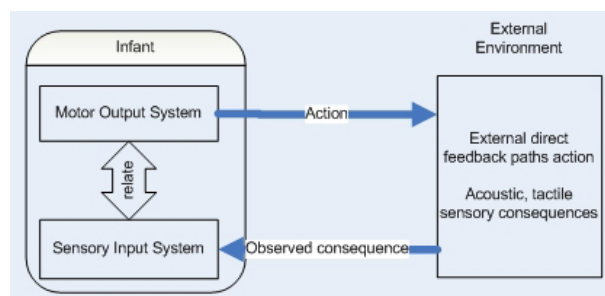

Figure 2: Signal flow interactions showing the body-external sensory consequences of his action.

2.5. Reinforcement feedback from a learned speaker

We argue that an infant will not be able to learning to speak simply by acoustic matching his speech with that of an adult's,

as suggested in some accounts of infant speech acquisition.. Firstly the vocal apparatus of an infant differs considerable from that of an adult's, and consequently linguistically equivalent sounds produced by these two systems will differ significantly in their acoustic properties. This means that a simple naive matching on the basis of spectral characteristics will not be adequate to evaluate similarity. At the very least we would need some kind of normalization. Although there is some evidence to suggest that infants can develop equivalence classes between utterances from different speakers [14], the problem we face here is not one of discrimination between different acoustic forms, but rather inferring a motor action from an acoustic form, The ability to discriminate acoustic input is not necessarily informative to the action that needs to be taken to generate it.
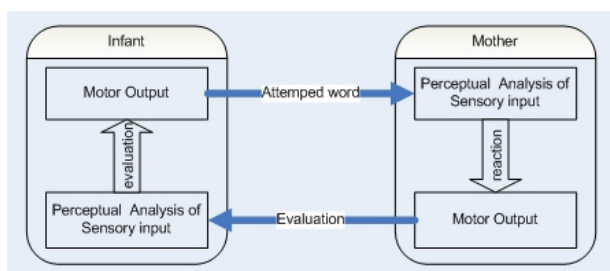

Figure 3: Signal flow interactions of an infant t receiving reinforcement feedback from his learned mother.

The comparison needed is defined at a linguistic level, and whether or not a young infant will initially have the ability to interpret speech in this way, his learned mother certainly will have it. As such her reaction to this speech provides a strong source of linguistic evaluation of his speech production. A major feedback path of reward thus becomes available to the infant when the environment includes his mother. There are essentially two ways in which she can aid speech development in the infant. Firstly she can evaluate the infant's speech and respond to sounds, and reward them, for example by giving attention in terms of a smile or a vocal response. This signal flow pathway is shown in figure 3. In this way she can reinforce gestures she considers interesting at his current point in development, and consequently guide his production of utterances towards L1. Evidence for this being the dominant speech production reinforcement path shows up in the observation that if an infant's mother makes no attempt to correct him, he will be quite happy producing his version of a given utterance even if it contains substantial errors, provided when he does so it elicits the desired response. This behavior is indeed observed in young infants.

2.6. Defining equivalence using the mother's feedback

Another important interaction that occurs between the infant and his mother is when she reformulates his speech, that is, when she 'imitates' the infant [15], as shown by the signal flow path in figure 4. The effectiveness of this reinforcement path for learning vowel quality has been demonstrated by Yoshikawa [16]. Provided the infant has understood that her utterance is intended to be equivalent to his (or because of repeated contingency), this enables him to associate an adult speech sound to his gestural formulation. This solves the correspondence problem and associates his motor action with

the linguistically equivalent adult form. This interaction path becomes even more effective when the infant is exposed to a wider range of learned speakers, as supported by observations, because they will have less ability, or indeed interest, in interpreting utterances produced by the infant if they differ significantly from the acceptable form in the ambient language.
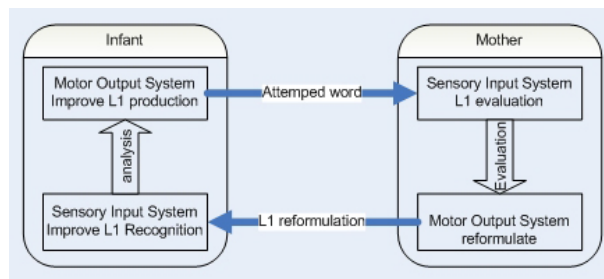

Figure 4: Signal flow interactions of an infant in his environment receiving reformulation from his learned mother.

2.7. Reinforcement learning

In terms of building a computational model of speech acquisition, how we use reward to learn motor patterns is an important issue. Learning algorithms generally involve making small changes in the parameters P and observing the effect in reward R. The parameters are then updated so as to increase reward. How exactly changes in the parameters are made depend upon the algorithm, as well as on the sub-structure of the parameter space, and so on. If a speech utterance contains several sub-actions, the task then also involves the classical credit assignment problem, that is, how much credit (or punishment) should each sub-action receive. Naive schemes may simply perform a random search of the parameter space, which totally ignore any structure between action and reward, whereas more sophisticated ones may use gradient information of the reward surface as well as any hierarchical structure. In general there will be local maxima in the reward surface, which may cause gradient decent techniques to become trapped in non-optimal solutions and there may also be many solutions with similar reward. In terms of movements patterns of the vocal apparatus, some of these local maxima in reward may correspond to different, but useful, speech sounds.

2.8. Motor pattern generation

The form of the motor pattern generator will affect how speech develops and how complex utterances can be constructed from smaller ones. If we wish to model how an infant's speech develops from the first sounds he makes, via the generation of the first simple words, through to the development of complex strings of words and account for the systematic errors and generalizations that occur, we must ensure that our model for the pattern generator has a similar computational structure to that of a real infant. The form of the neural representation of movement is still a matter of considerable controversy [17], although there is evidence that a represrentation in terms of target positions is a useful one. We define a basic motor pattern in terms of starting and

ending articulator targets positions and associated transition duration (although this was fixed to 500ms in our experiments). In this preliminary work we simply implement the mapping M, which generates paths from the sequence of targets, by interpolating between them using critically damped trajectories, as adopted my Markey [11]. This formulation is equivalent to using a low level controller to move the articulators from their starting positions towards the target positions without overshoot, taking into account the dynamics of the vocal tract. It also assumes that the movements are not perturbed from their trajectories by noise, or by internal or external obstructions.

2.9. Hierarchical utterance generation

Adult speech utterances exhibit sub-structure, with sub-patterns of articulator movements repeating within different words. It has been observed that in infant speech production development up to the so called 50 word stage, words appear to be treated as whole units without noticeable sub-structure. However after this stage has been reached sub-word structure becomes apparent in the speech. At this stage, the infant's productions also generally still differ from adult word form, even though more than half of them can be identified by adults. Usually after this stage is passed, the addition of the next new words results in a degradation of performance, which then slowly recovers, followed by a rapid increase in vocabulary over a relatively short time.

2.10. Using sub-structure to accelerate word learning

Sub-structure provides a means to generate more complex speech output by repeating previously learned sounds in different combinations. If basic motor patterns are used as building blocks for more complicated ones, the latter can be found without the computational penalty (which would arise because such a pattern would have more degrees of freedom than a basic one) that would be incurred by searching for such a pattern from scratch. Such reformulated babble still needs to be rewarded by the mother to signify its value, just as basic motor patterns need to be reinforced for linguistic significance. The point is, however, that the starting point for these complex motor patterns is essentially available for free.

As we discussed before, perturbing actions and rewarding ones that generate linguistically significant output events may work well as a mechanism to train the first stages of production, and we suggest that a stage of development it is the most effective mechanism available. However, it would be much more efficient if a child could listen to his mother and then directly imitate what is heard using a mechanism to directly map sensory input onto speech production. Elsewhere we argue against the possibility that a child can initially imitate speech sounds [9]. However the emergence of simple sub-patterns in production permits the development of the necessary associations between the infant's and mother's speech, driven by her reformulations. As soon as the child becomes aware of sub-structure in his production, he can learn the equivalence between his sub-patterns of production to those of his mother's and thus has learned how to imitate! Using this new ability, the mother's speech can be analyzed and the child equivalent form can be activated using these associations. The emergence of such an ability would clearly account for the rapid increase in acquisition of vocabulary that

is observed as soon as sub-structure is employed in production.

3. Computational Implementation

3.1. Introduction to the computational model

We now discuss an implementation of a computational model based of the previous ideas. We model the vocal apparatus, the motor system that drives it as well as a simple perceptive evaluation, and implement simple motor pattern learning mechanism based on optimizing the reward generated by the production of output. Initially the model runs stand-alone, then later a human subject evaluates its speech production. The articulator synthesiser was implemented in C and all other analyses were implemented and run using Matlab.

3.2. Model of the vocal apparatus

To model the vocal apparatus of an infant, we physically scaled down the dimensional parameters of a 9-parameter articulator synthesizer based on the work of Maeda [18]. To limit computational requirements, we adopt a speech sampling rate of 8kHz. We model that fact that articulator control increases slowly after birth by initially only permitting jaw and lip movements and only enable tongue movements after initial salience driven reward optimization has learned to control the simplified model. The parameters in the two stages are shown in table 1. We generate proprioceptive feedback signals from lip and tongue contact by observing the place and times at which the synthesizer's vocal tract cross-sectional area reaches zero.

Young Infant	Older Child
Reduced articulator control Clamping some parameters to zero	Full articulator control (no nose!)
p1 Jaw position	p1 Jaw position
p2 Tongue dorsum position	p2 Tongue dorsum position
p3 Tongue dorsum shape	p3 Tongue dorsum shape
p4 Tongue apex position	p4 Tongue apex position
p5 Lip height (aperture)	p5 Lip height (aperture)
p6 Lip protrusion	p6 Lip protrusion
p7 Larynx height	p7 Larynx height
p8 Voicing	p8 Voicing
p9 Fundamental frequency	p9 Fundamental frequency

Table 1: Maeda Parameters

3.3. Model of the motor system

As we mentioned above, a motor pattern is defined in terms of start and end target articulator positions and transition duration, and the articulators are moved along critically damped trajectories between these positions. Such a motor pattern can be thought of as basic representation of speech at the level of the syllable. By repeating such motor patterns or combinations of them, speech-like utterances and babble can be generated. Using this model, learning basic motor patterns involves finding appropriate articulator targets and durations.

Leaning to generate more complex utterances (e.g. words) involves finding appropriate sequences of motor patterns.

3.4. Model of the perceptive system

In this preliminary work, we limit the perceptual system to the computation of salience of the sensory consequences of the model's own actions, and a direct external evaluation path from a learned listener. Input speech, generated by the synthesiser, is first filtered using a 800Hz 2pole LPF and low frequency power is then computed within a 125ms time window. Spectral change is computed by summing the output a differenced narrow-band spectrogram over all frequencies and over a 125ms time window. Tongue and lip contact also contribute to sensory salience, which is then considered to be an important event. We compute reward as the average of the sensory salience components minus the contribution due to effort. In this simple implementation, the latter is just taken as the degree of voicing in the synthesizer, although future work this will also include a contribution from the squared speed of the articulators. Table 2 shows the components that were averaged to calculate reward.

3.5. Learning based on salience

Saliency components Signals are scaled to range 0-1 and reward is given as mean of the individual components	Contribution
Low Frequency Power after 800Hz 2pole LPF	+ve
Spectral change differenced narrow-band Spectrogram	+ve
Lip Contact, found from cross sectional area,	+ve
Tongue Contact, found when cross sectional area goes to zero.	+ve
Effort represented by degree of voicing	-ve

Table 2: Contributions to Internal Reward

In the first experiments, the vocal tract synthesizer was run with the limited infant parameter set, i.e. with no control over tongue position, which was instead kept in a neural position. To search for simple speech-like sounds on the basis of their sensory salience, a set of basic motor patterns were generated randomly and an optimization algorithm was run on each of them to find a set of solutions that generate salient output. For each initial motor pattern, the optimization involved iteratively applying a constrained gradient decent procedure, which used the following steps. Firstly, one seconds of output was generated from the vocal tract synthesizer using the current estimate of the motor pattern. The sensory consequences were then analyzed and via salience and effort computations an associated reward value was found. The optimization algorithm used this reward to update the motor pattern. This was run for 30 cycles, by which time a useful motor pattern had generally been found or only a poor solution had been found which did not improve by substantially increasing the number of optimization cycles (because it was probably stuck in a local reward maximum). Each final motor pattern was then recorded together with its reward value. To reduce computation time in the simulation, the exploration was limited to one hundred different articulator configurations. The optimization of articulator movements based on the salience lead to the generation of useful baby-like speech sound, and for evaluation purposes synthesizer output was generated by repeating each of these motor patterns twice (see discussion). These motor patterns were then copied and the optimization procedure was run again in exactly the same fashion, but this time all vocal tract parameters were free for modification by the algorithm, i.e. tongue movement was now permitted, representing a young child with more vocal tract control. (We noted that starting the full parameter control case from an initially random start significantly reduced the number of motor patterns that generated speech-like sounds, because the optimization then had a larger number of degrees of freedom from the start and consequently it was harder to find solutions that produced useful sounds). The further optimization resulted in some motor pattern with tongue movement, giving rise to baby-like speech sound with changes in vowel quality. Again, for evaluation purposes, synthesizer output was generated by repeating each of these motor patterns twice (once again, see discussion).

3.6. Evaluation using the infant's mother

As previously discussed, optimization of motor speech patterns based on the reinforcement of those that generate useful sounds from a learned adult can greatly assists the development. To avoid modeling perception, we used a simple feedback evaluation scheme to enable an adult listener reward or punish the utterances generated by the model. This involved him listening to all the speech like utterance generated by the different motor patterns, asking him to ranking each one as good, neutral or bad. This procedure was carried out on the two groups of patterns, once for those found using the infant's vocal tract (no tongue movement) and once for those found running an additional stage of optimization using the child's vocal tract (with tongue movement). By ranking them in terms of their perceived usefulness to the listener, this selection procedure rapidly weeds out the motor patterns that do not sound like speech. As a final experimental the 8 highest ranked basic motor patterns were selected (to limit their number) and all possible combinations of these motor patterns were again evaluated by a listener, resulting in the discovery of several quite realistic speech-like word forms.

4. Discussion

4.1. Preliminary experimental results

We recommend that readers visit the supplementary information to this paper available on the internet at http://www.ianhoward.info/Specom2007.htm. This contains more information and block diagrams to help explain our computational model. It provides examples of the speech utterances learned by the computational model in the form of .wav format audio files. This first includes output from good reinforced motor patterns for the vocal tract synthesizer trained as an infant (partial parameter control) and as a child (full parameter control). Output for good reinforced reduplicated babble built from the best of these motor patterns is also demonstrated.

Optimization of articulator movements based on the infant's own salience evaluation leads to the generation of useful speech-like sounds. Starting with a limited parameter set,

simple speech-like sounds were quickly found. By subsequently increasing the degrees of freedom of the vocal tract, sound variety was increased (in particular the vowel qualities). Using evaluation from a learned listener, these simple sounds were then rewarded or punished, leading to a collection of speech-like sounds. By recombining these basic motor patterns together in pairs, and again using listener evaluation, realistic reduplicated infant babble was generated.

5. Conclusions

5.1. Summary

We demonstrated that our model learns to speak only on the basis of reinforcement feedback and without using acoustic imitation. The reward feedback was not informative about what the motor patterns should be like and only providing a scalar evaluation and this was sufficient to drive the learning. The results also show how important the biophysical form of the vocal tract is to speech development and that it was relatively easy to find for motor patterns that give rise to simple speech-like sounds just by optimizing the motor patterns on the basis of their sensory salience. The interaction with a learned adult very quickly weeded out non speech-like sounds, which could not be distinguished on the basis of the model's reward based salience measure. Combinations of the basic patterns followed by further listener evaluation could then be used to generate realistic reduplicated babble.

5.2. Future work

The current Maeda synthesizer implementation lacks control over the nasal cavity. Since some of the first important speech utterances generated by real infants are the CVs such as /ba/ and /ma/, to make our infant's speech output more realistic, we will soon use a synthesizer that includes nasal control.

We are also currently implementing a mechanism to take advantage of reformulation. We will provide a list of simple words the synthesizer can generate and a subject will be asked to try to reformulate the model's output if it comes close to one of them. An absence of response from the subject will be used to indicate no or negative reinforcement. The reformulation will be used to train a simple DTW recognizer and associate it bidirectional with the model's corresponding motor pattern. The subject's future acoustic realization of the same word should then trigger the model to generate its version of the word.

6. Acknowledgements

7. References

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

[2]   Kuhl, P. K. (2000) 'A new view of language acquisition.' Proc Natl Acad Sci USA 97.22: 11850-11857.

[3]   Westermann, G. and Miranda, E. (2004) A new model of sensorimotor coupling in the development of speech. Brain and Language, 89, 393-400.

[4]   Bailly, G. 1997. ''Learning to speak. Sensori-motor control of speech movements,'' Speech Commun. 22, 251–267.

[5]   Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006) 'Neural modeling and imaging of the cortical interactions underlying syllable production.' Brain and Language 96: 280-301

[6]   Kuhl, P.K., Meltzoff, A.N. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. Journal of the Acoustical Society of America 100 (4), 2425-2438.

[7]   Howard, I. S. & Huckvale, M. A., ''Learning to Control an Articulator Synthesizer by Imitating Real Speech'', ZASPIL, Franco-German Summer School, Lubmin, Germany 2004.

[8]   Howard, I.S & Huckvale, M.A., "Training a Vocal Tract Synthesizer to Imitate Speech using Distal Supervised Learning ",, Specom 2005, Patras, Greece.

[9]   Messum, P.R. (2007) "The Role of Imitation in Learning to Pronounce, PhD Thesis, University College London

[10] Menn, L., K. L. Markey, M. Mozer, and C. Lewis. 1993. Connectionist modeling and the microstructure of phonological development: a progress report. In Developmental Neurocognition: Speech and Face Processing in the First Year of Life edited by de Boysson-Bardies, B. and et al., 421-433 (Dordrecht: Kluwer).

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

[12] Blakemore SJ, Wolpert DM & Frith CD (1998) "Central cancellation of self-produced tickle sensation" Nature Neuroscience 1(7):635-640

[13] Shergill SS, Bays PM, Frith CD & Wolpert DM (2003) "Two eyes for an eye: The neuroscience of force escalation", Science 301:187

[14] Kuhl, P. K. 1991. Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In: Brauth, S.E., Hall, W.S., Dooling, R.J. (eds) Plasticity of Development 79 Cambridge MA: MIT Press.

[15] Pawlby, S. J. 1977. Imitative interaction. In: Schaffer, H. R. (ed) *Studies in Mother-Infant Interaction*, 203-223 London: Academic Press.

[16] Yoshikawa, Y., Asada, M., Hosoda, K., Koga, J. 2003. A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science* 14 (4), 245-258

[17] Todorov E , "Direct cortical control of muscle activation in voluntary arm movements: a model", *Nature Neuroscience* 3(4): 391-398 (2000).

[18] Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In Hardcastle W.J. and A. Marchal (eds.) Speech production and speech modelling. Kluwer Academic Publishers, Boston. p.131-149