ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Brain and Language xxx (2004) xxx–xxx

Brain and Language

www.elsevier.com/locate/b&l

A new model of sensorimotor coupling in the development of speech

Gert Westermann[a,*] and Eduardo Reck Miranda[b]

[a] Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

[b] Computer Music Research and Centre for Theoretical and Computational Neuroscience, Faculty of Technology, University of Plymouth Drake Circus, Plymouth PL4 8AA, UK

Accepted 20 August 2003

Abstract

We present a computational model that learns a coupling between motor parameters and their sensory consequences in vocal production during a babbling phase. Based on the coupling, preferred motor parameters and prototypically perceived sounds develop concurrently. Exposure to an ambient language modifies perception to coincide with the sounds from the language. The model develops motor mirror neurons that are active when an external sound is perceived. An extension to visual mirror neurons for oral gestures is suggested.

1. Introduction

In recent years there has been mounting evidence that the pre-linguistic period plays a major role in the development of phonology. It is often hypothesized that the first speech-like articulations and the babbling phase between 5 and 10 months of age allow infants to develop a link between articulatory settings and the resulting auditory consequences. This link forms the basis for the development of the phonetic inventory and the adaptation to the ambient language by exposure to other speakers.

Early theories of speech perception have argued for an innately specified, evolved link between perception and action and a representation of perceived speech in terms of their associated motor articulatory categories (Liberman & Mattingly, 1985; Liberman, Harris, Hoffman, & Griffith, 1957). More recent evidence suggests that this link is learned rather than innate, focusing on evidence for the role of babbling in normal speech development. It has been found that deaf infants do not babble normally, indicating the role of audition for normal babbling (Oller & Eilers, 1988), and that they subsequently often do not develop intelligible speech (Wallace, Menn, & Yoshinaga-Itano, 1998). Infants that were prevented from normal babbling due to tracheotomy likewise show abnormal patterns of vocal expressions that persist (Bleile, Stark, & McGowan, 1993; Locke & Pearson, 1990). It has also been suggested that motor performance has an effect on auditory perception (Locke, 1986; Vihman & Nakai, 2003), suggesting that the link from motor to auditory representations is likewise learned.

The coupling between auditory perception and motor production is closely linked to the ability of infants to imitate speech sounds. Vocalizations in babbling move from a language universal pattern to one that is specific to the ambient language of the child (Boysson-Bardies, Halle, Sagart, & Durand, 1989; Boysson-Bardies & Vihman, 1991). The ability to imitate vowel sounds seems to emerge between 12 and 20 weeks of age (Kuhl & Meltzoff, 1996). The pre-linguistic ability to produce sounds of the ambient language is seen as an important step in the development of a phonological inventory, first words and more complex linguistic structures (e.g., Vihman, 2002). A proposed mechanism for this development is the Articulatory Filter Hypothesis (Vihman, 1991, Vihman, 1993) which suggests that the experience of frequently self-producing CV syllables sensitizes infants to similar patterns in their ambient language, making these forms more salient as potential building blocks for first words. However, the more precise mechanisms of the learned sensorimotor linkage and of

* Corresponding author. Current address: Department of Psychology, Oxford Brookes University, Oxford OX3 OBP, UK.

E-mail address: gwestermann@brookes.ac.uk (G. Westermann).

0093-934X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.

doi:10.1016/S0093-934X(03)00345-6

the resulting changes in speech perception and production are unknown.

In this paper we suggest such mechanisms in a computational model of the development of a coupling between perception and action in the production of vowel sounds. Based on this coupling, preferred articulatory parameters and auditory prototypes develop concurrently, and the coupling enables the model to imitate heard sounds. Exposure to an ambient language alters the representations in the model towards the vowels from that language.

The model suggests a simple mechanism as the basis for phonological development in the infant babbling phase: a coupling between the representations of articulatory parameters and auditory perception that develops in an experience-dependent way and in this process alters both perceptual and motor representations.

Development in the model leads to the emergence of mirror neurons for acoustic stimuli, and an extension to visual mirror neurons that become activated by observing oral gestures is proposed. Mirror neurons have, for example, been found in the pre-motor cortex of the monkey (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). These are neurons that fire when the monkey performs an action, but also when he observes a similar action performed by another monkey or by someone else. Mirror neurons also exist in humans, where the same premotor regions respond to grasping movements by the subject and the observation of such movements in others (Rizzolatti, Fadiga, Gallesea, & Fogassi, 1996). They have been argued to play a role in the evolution (Rizzolatti & Arbib, 1998) and the development (Vihman, 2002) of language, although the nature of their role has been contested (Hurford, in press). While speech mirror neurons have not yet been found, our model, and the assumption of unified processing mechanisms in the cortex, suggest their existence.

In the rest of the paper, we first describe the model. Then, we explore the development of preferred motor patterns, perceptual prototypes and auditory mirror neurons in a babbling phase, and we show how the model can imitate heard sounds based on the developed sensorimotor coupling. Next, we demonstrate the adaptation of the model to an ambient language. An extension of the model to incorporate visual stimuli is then suggested and we show how this extension could lead to the emergence of visual mirror neurons. Finally, we discuss the implications of the model.

2. The model

This section describes the architecture and functioning of the sensorimotor integration model. A more detailed description, including the mathematical equations and parameter settings, is given in (Westermann & Miranda, 2002).

The basic idea of the model is this: the model consists of two domain maps, one for auditory stimuli and the other for motor commands (Fig. 1A). Connections between these two maps develop in an experience-dependent way based on the activity on each map. The model can either *babble* by generating a set of motor parameters and 'listening' to the resulting sound—motor command and resulting sound then form the inputs to the corresponding maps—or *listen* to external sounds, in which case only the auditory map receives external input. The development of connections between the maps leads to changes of the representations in each map, explaining the changes in perception and production during infant speech development. In this model we restrict sounds to vowels because they have a static representation. The motor parameters used in the model represent the positions of different articulators (see below), and vowel sounds are represented by their first two formant values.

Each neural map consists of a number of randomly placed units. The units act as receptive fields with a Gaussian (bell-shaped) activation function. Units that are close to an input signal (a large circle in Fig. 1A) become highly activated (lighter color of units in the figure), and activation decreases with distance to the

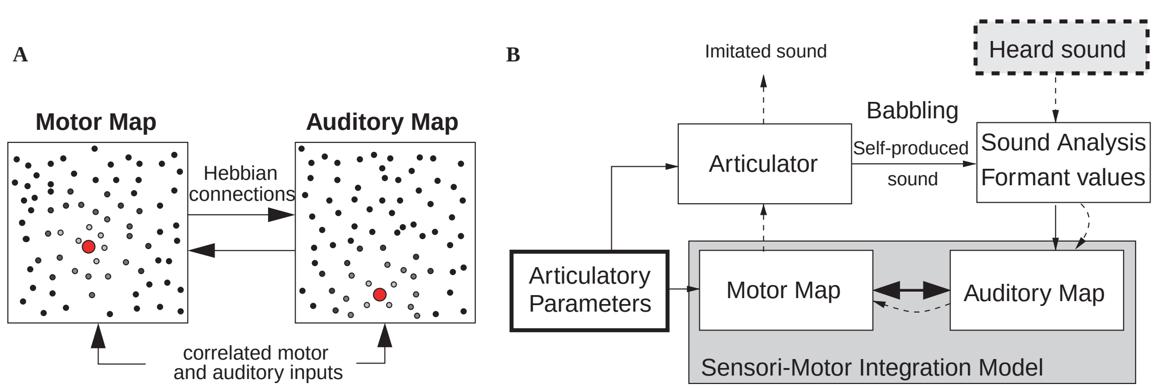

Fig. 1. The basic architecture of the sensorimotor integration model (A) and its embedding in the experimental setup (B) to learn the coupling between articulatory parameters and self-generated sounds through babbling. Once learned, the system imitates heard sounds (dashed lines).

signal. Such receptive fields are ubiquitous in the cortex. Typically each input activates a number of units. The response to an input, that is, how the map perceives the input, is computed as a population code: the perceived signal is the vector sum of the positions of all active units, weighted by their activation values. Such population codes have been found to exist in many areas of the cortex, for example in the monkey motor cortex (Georgopoulos, Kettner, & Schwartz, 1988).

When an input signal is presented to the map, the population-coded response corresponds to this actual input. The principle of the model is, however, to integrate both maps, and this integration alters the responses of the maps, that is, the motor parameters and perceived sounds. The integration of the maps is achieved with Hebbian connections between the units on each map. During the simulated babbling phase, a motor setting and its resulting sound are presented to the maps simultaneously. When the activation of a unit on, for example, the motor map consistently co-varies with the activation of another unit on the auditory map, connections develop between these two units. The more reliable the co-variance, the stronger become the connections. As a consequence a unit is not only activated by the input from its own domain, but it receives additional activation from the other domain via the developed Hebbian connections.

The extra activation shifts the population-coded response toward those units that are most highly activated by the external stimulus and additionally receive the highest activation from the other map. Because units that co-vary reliably with units from the other map develop the strongest connections and thus receive the highest additional input, responses will be mainly shifted toward these units. In this way, responses become preferred if they consistently co-occur with the same responses in the other domain. That is, the model develops preferred responses for highly correlated motor–sound pairs.

The following sections describe experiments with the model. The setup of the experiment is shown in Fig. 1B. In the babbling phase, the model creates random articulatory parameters that form the inputs to the motor map and are used by a physical model of the vocal system (Boersma & Weenink, 1996, see (Miranda, 2002) and (Westermann & Miranda, 2002) for a detailed description) to produce the sounds. These sounds are analyzed in terms of their first two formant values, which then form the inputs to the auditory map. During this phase, weights between the maps develop as described above.

3. Imitation based on production

Given the evidence that the coupling between perception and production is learned and forms the basis for the imitation of speech sounds in infants, in a first experiments we explored how the model could account for these results and how the sensorimotor coupling would alter both perception and production during a babbling phase. For this purpose, we created a set of motor commands by continuously varying two motor parameters, jaw opening and the position of the styloglossus muscle (a muscle at the back of the tongue that controls tongue elevation), at 18 steps each. This procedure resulted in a set of 324 motor parameter sets for which the 324 vowel sounds were generated with the vocal synthesizer.

The model babbled by randomly choosing a motor parameter set with its associated sound and using each as input to its corresponding map. The babbling phase consisted of a single run through all 324 articulations, and weights between the maps were updated after each articulation. Choice of model parameters affected the precise results while the qualitative results were robust. Here we report results from individual runs of the simulation.

The responses of the model are shown in Fig. 2. Initially, before Hebbian weights have developed, the population-decoded responses faithfully represent the continuously varied motor parameters and their associated sounds (Fig. 2A). After the training of the model (Fig. 2B), both motor and auditory responses have

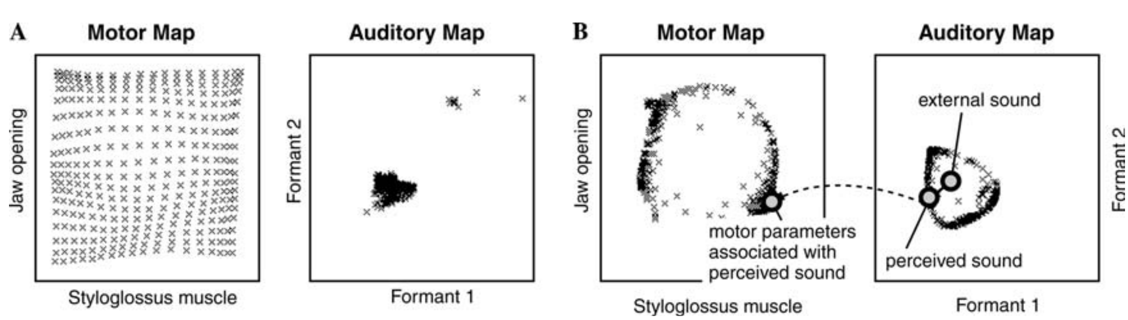

Fig. 2. Response of the model to motor and sound parameters before (A) and after (B) training of the Hebbian connections. After training an external sound is perceived in terms of the developed auditory prototypes and mapped onto the corresponding articulatory parameters.

changed significantly. On both maps, preferred responses have developed as dense clusters.

What are these preferred responses? The mapping between articulatory parameters and the resulting sounds is highly non-linear: small changes in articulator settings can lead to large changes in sounds, and near identical sounds can be produced by different articulator settings. The highest correlation between motor parameters and sounds occurs in the areas where this mapping is more linear. When small changes of articulatory parameters only lead to small changes in the resulting sound, the same units on the maps are involved in the representations of many of these motor/sound pairs, and strong inter-domain connections develop. As a consequence, preferred responses emerge for the linear regions of the motor/sound mapping, that is, for sounds that can be reliably produced.

After the model has developed the sensorimotor coupling, it can utilize this coupling to imitate heard sounds (Fig. 2B). A heard sound evokes a response on the auditory map, and through the Hebbian connections, an associated motor parameter set is activated that can then be used to produce the sound. Imitation here is not the accurate re-play of the heard sound, but instead an interpretation of this sound within the developed sensorimotor framework.

4. Adaptation to the language environment

Studies of infant babbling have shown that babbling progresses from a language universal phase to one that is specific to the ambient language of the infant (Boysson-Bardies et al., 1989; Boysson-Bardies & Vihman, 1991). This development has been linked to an increase in vocal imitation between 12 and 20 weeks of age (Kuhl & Meltzoff, 1996). However, it is unclear how perception of a non-self-produced sound can lead to a preferred selection of the articulatory parameters associated with that sound. While the Motor Theory of Speech (Liberman et al., 1957; Liberman & Mattingly, 1985) assumes an innate link between perceptual and production parameters, a developmental account is given by the Articulatory Filter Hypothesis (Vihman, 1993). According to this hypothesis, experiences with speech patterns in the ambient language highlight the corresponding motor–sound pairs that have been learned through babbling. (Vihman, 2002) argued that the best-represented patterns of adult speech, that is, those adult patterns that most closely resemble the child's most typical production patterns, form the basis for the first words.

A second experiment modelled the adaptation of the model to its external environment. This experiment investigated how the presence of sounds that are not produced by the model itself shape its perceptual and motor properties. For this purpose a more sophisticated set of motor parameters was created by continuously varying six motor parameters: the styloglossus muscle to control backward and upward tongue movement; the hypoglossus muscle to control backward and downward tongue movement; the levator palatini muscle to control the raising of the velum; the cricothyroid muscle for vocal chord stretching, the interarytenoid muscle for vocal chord adduction, and lung pressure. By this method, 1876 different motor parameter sets were generated, and for each set a sound was produced with the vocal synthesizer. Sounds were again analyzed in terms of their first two formant values.

Two different ambient language environments were simulated: a French environment consisted of 11 vowels segmented from the speech of a native French speaker and a German environment comprised 15 vowels from a native German speaker. All vowels were encoded in terms of their first two formant values. While self-generated sounds existed in motor parameter/formant pairs, the ambient sounds had no equivalent motor parameter set.

The model was trained as in the previous experiment. Motor parameter sets were randomly chosen and presented to the model together with their corresponding sound. Training was either babbling only (20,000 articulations), or babbling that was randomly interspersed with either French or German vowels (20,000 instances with 10% self-generated articulations and 90% vowels from the ambient language).

Fig. 3 shows the responses of the auditory map before (A) and after (B) training the model *without* exposure to any of the external vowels. Like in the previous experiment, before training (A), the responses of the model faithfully represent the actual sounds generated by the vocal synthesizer. After training (B), the model has again formed dense clusters of prototypical sounds. French and German vowels are indicated by squares and circles, respectively. Without exposure to these vowels the development of prototypes depends entirely on the non-linearities between the articulatory settings and the resulting sounds. It is remarkable, however, that many of the developed prototypes lie very close to the actual French and German vowels.

This surprising result suggests that in language evolution, vowel inventories might have developed as the outcome of perceptual and articulatory constraints and the developing sensorimotor coupling in infancy. A similar hypothesis has been put forward by Lindblom (2000). Lindblom considered ease of articulation, or "minimum energy expenditure" as the dominating constraint leading to the emergence of adult phonological inventories. Since in the present model preferred motor/sound settings develop for sounds that can be produced reliably (see above), these results suggest that stability, or "reliability of production," might be another important constraint. The phonological inventory

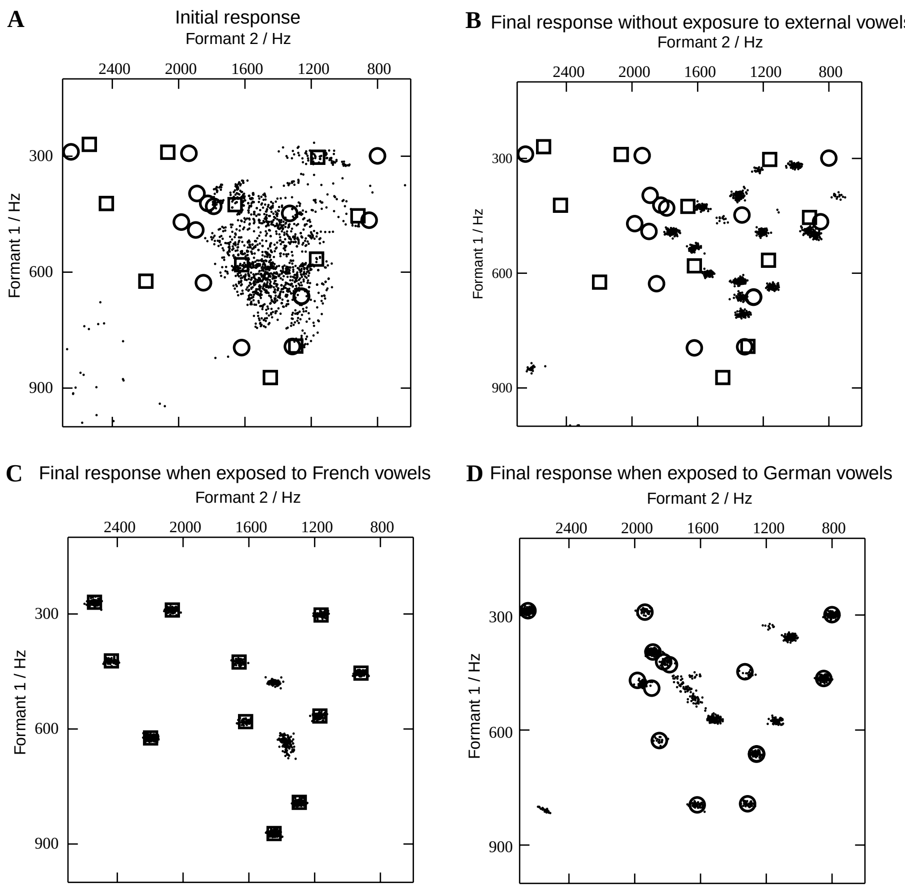

Fig. 3. Responses of the model to self-generated sounds under different training conditions. French vowels are indicated by squares, German vowels by circles.

that emerges from these production-oriented constraints is additionally subject to perception-oriented and cultural constraints such as discriminability (see also Lindblom, 1986). In Fig. 3 this is suggested by the fact that the space covered by the French and German phonemes is more stretched out than that covered by the internally generated prototypes. The additional constraints lead to a modified phonology that reflects a trade-off between productive and perceptual optimality.

Although the model suggests that the preferred selection of stable motor parameters is an emergent property of the sensorimotor coupling, an emergence of this preference through evolutionary pressure seems also plausible. An infant with an as yet unskilled articulatory system can produce reliably only those sounds that stay relatively stable under varying articulatory settings. Likewise, from the perspective of communication, sounds that are stable in this way would form more reliable building blocks for the phonological inventory of a language.

Figs. 3C and D show the responses of the model after babbling when it is exposed to French and German vowels, respectively. Now the perceptual responses (and through the sensorimotor coupling, the preferred motor parameters) to the majority of sounds correspond to the sounds in the ambient language. A small number of perceived sounds do not fall onto the ambient sounds, indicating that the ability to distinguish vowel sounds that are not from the ambient language has not been completely lost. This is because in the model, the final perceptual map develops based on the interactions between self-generated prototypes and ambient vowels, and some of the self-generated prototypes can be preserved (or only slightly altered) by exposure to the ambient language.

How does the exposure to an external sound that does not have an associated motor parameter, alter the perception (and production) preference of the model? The responsible mechanism in our model closely corresponds to the Articulatory Filter Hypothesis (Vihman, 1991; Vihman, 1993). By producing babbling sounds, the model starts to develop connections between the motor and the auditory map based on the co-variance between articulatory parameters and their auditory consequences. A subsequent external sound directly activates only the auditory map. However, some motor units receive activation through the developing Hebbian connections from the active auditory units. As a consequence, auditory and motor units are active simultaneously and the Hebbian weights are strengthened. In this way sounds from the ambient language reinforce selectively these sounds from the infant's babbling inventory.

This mechanism suggests that the adaptation of infant babbling to the ambient language might not depend on the explicit imitation of sounds by the infant (Kuhl, 2000), but that instead imitation is made possible by the reinforcement-based adaptation. It also suggests how mirror neurons develop in the motor area that respond when an external sound is heard. These are the same units that would be active when the model produces the sound by itself. It should be noted that mirror neurons in the present model are an emergent property of the sensorimotor coupling and they do not per se represent a mechanism for the imitation of sounds. Instead, they reflect the multi-modal representation of speech sounds.

5. Visual mirror neurons for speech

Many experiments have shown that visual information can enhance the understanding of speech, suggesting an integration of the visual with the auditory signal in this task (see e.g., Massaro, 1998, for an overview). This integration is already present in young infants between 18 and 20 weeks of age (Kuhl & Meltzoff, 1982). In their study, infants were presented with two side-by-side images of a female speaker producing the vowels /a/ and /i/, respectively. The infants were familiarized on the films without sound. In the test phase the same sound was added to each film. It was found that the infants looked significantly longer at the face where sound and lip movement matched, indicating that they were sensitive to the relation between lip movements and the resulting sound.

A simple (not yet implemented) extension to the sensorimotor model could account for this effect (Fig. 4). When a visual map is added to the model that responds differentially to different lip movements, the visual stimulus and the auditory stimulus co-vary (a). Based on previous babbling, the auditory stimulus activates the associated motor parameter set on the motor map (b). As a consequence, visual and motor activations now co-vary as well (c), and Hebbian connections between the visual and articulatory maps can develop. This mechanisms will lead to two sources of input to the articulatory map when a speaker is seen and heard. Matching auditory and visual signals will then activate the same group of units on the articulatory map, whereas a mismatch will lead to different groups of units being activated, which could explain the preference of infants to matching stimuli.

A model that learns the match between auditory and visual stimuli in this way develops a tri-modal representation of speech, with both auditory and visual stimuli alone resulting in activation on the articulatory map. In this way, articulatory neurons act as auditory as well as visual mirror neurons and can become the basis of the visual imitation of oral gestures. It has been hypothesized that visual imitation plays an important role in the development of, e.g., labials (McCune & Vihman, 2002).

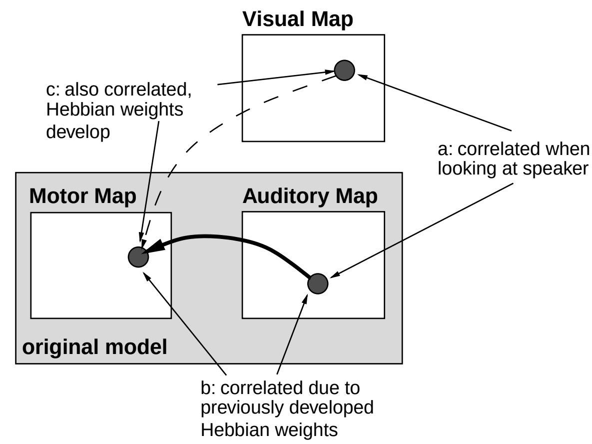

Fig. 4. Development of visual mirror neurons in an extension of the model.

6. Discussion

The model presented here suggests a simple mechanism—the coupling between the representations of articulatory parameters and auditory perception—as the basis for phonological development in the infant babbling phase. The model is based on neurobiologically plausible principles, namely, neural receptive fields, Hebbian connections that develop in an experience-dependent way, and population-coded responses to multidimensional stimuli.

A significant implication of the model is that motor practice should have an effect on auditory perception. Such a link has been suggested by Locke (1986) and has recently been shown experimentally for monolingual English and Welsh children at 12 months of age (Vihman & Nakai, 2003). In their study, two consonants in each language were identified that had roughly the same frequency in infant-directed speech, but that differed in the production frequency of the infants. The idea was that any perceptual differences between the two phonemes in the infants would have to be attributed to their differing production frequencies. When tested on both consonants, the infants showed a novelty preference for their less-favored consonant. An explanation of this result in terms of the present model would be that a sound that has been produced often results in a tight sensorimotor coupling with a high activation of the perceptual prototype, whereas a less frequently produced sound with a weaker coupling would result in a smaller activation of the perceptual map, indicating less familiarity.

This view, and our model, also suggest that in the absence of normal babbling (e.g., due to tracheostomy) infants will not only develop abnormal production patterns but also abnormal perception. We are not aware of empirical investigations so far of this prediction.

The model can also give insight into why it is difficult to learn a new phonology when learning a second language. Speakers who learn a new language after a certain age usually retain an accent characteristic of their native language. Kuhl (2000) has argued that this inability to fully learn a new phonology might be a consequence of neural commitment to the original phonological map developed for the first language. While this view implies a loss of neural plasticity, a slightly different angle on this hypothesis suggested by the model is that sound differences that are important in one language might not be important in another and might not even be perceived as different (e.g., the /r/-/l/ distinction in Japanese, Strange & Dittmann, 1984). Neural adaptation for learning to produce two different sounds, however, can only occur if the two sounds are perceived as different: in the model, if both are mapped onto the same developed perceptual prototype, they share the same perceptuo-motor representation and there is no basis for their distinction. Thus, when new sounds are perceived as if they were known sounds, even with retained neural plasticity, there is no adaptation because there is nothing new to learn. This view suggests that the real problem of re-learning phonology might be perceptually based instead of neurally based, instead of ''neural commitment'' (loss of plasticity) there is ''perceptual commitment'' (altered perception leading to the loss of perceptual distinction) (see also Iverson et al., 2003).

Our model could be extended in several ways. A restriction of the current model is that it is concerned only with the acquisition of vowel sounds that can be represented statically. A more comprehensive model would have to include consonants that vary over time, and eventually, sequences of sounds as they occur in canonical babbling and beyond. Further, an important constraint in the development of vocalization is the anatomical development of the speech apparatus. On one hand it allows for a progressive increase in the vocal repertoire, on the other it requires a continuous, if gradual, re-mapping of the sensorimotor coupling. To incorporate this aspect of development into our model would require motor/sound pairs that change over time, simulating the changes in the articulator structure. In terms of theories of language evolution, the current model does not incorporate the ''ease-of-articulation'' constraint proposed by Lindblom (2000), because this constraint relies on preferences for certain articulator settings. In our simulations, however, production parameters were varied continuously and each parameter setting was equally likely. To account for the ease-of-articulation constraint, parameter selection could be biased, e.g., to avoid extreme positions of the articulators.

By restricting the model in these ways, however, we believe that we have been able to uncover several important principles of the development of vocalization and the ensuing sensorimotor coupling. The model provides a mechanistic account of previous theories of the development and evolution of phonetic inventories (Lindblom, 2000; Vihman, 1991,Vihman, 1993,Vihman, 2002), and at the same time it can serve as inspiration for further empirical work addressing the development of mirror neurons, the nature of preferred motor patterns and perceptual prototypes, and the development of multi-modal integration in speech perception.

Acknowledgments

The writing of this paper was supported by European Commission RTN Grant HPRN-CT-2000-00065 to Gert Westermann. We thank Marilyn Vihman and two anonymous reviewers for helpful comments on a draft of the manuscript.

References

Bleile, K., Stark, R., & McGowan, J. (1993). Speech development in a child after decannulation—further evidence that babbling facilitates later speech development. *Clinical Linguistics & Phonetics, 7*, 319–337.

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer*. Tech. Rep. 132, Institute of Phonetic Sciences of the University of Amsterdam.

Boysson-Bardies, B. d., Halle, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*, 1–17.

Boysson-Bardies, B. d., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling of infants according to target language. *Language, 67*, 297–319.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.

Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neural population. *Journal of Neuroscience, 8*, 2928–2937.

Hurford, J. R. (in press). Language beyond our grasp: What mirror neurons can, and cannot, do for language evolution. In K. Oller, U. Griebel, & K. Plunkett (Eds.), *The evolution of communicative systems: A comparative approach*. MIT Press.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47–B57.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, USA, 97*, 11850–11857.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

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*, 2425–2438.

Liberman, A., & Mattingly, I. (1985). The motor theory of speech-perception revised. *Cognition, 21*, 1–36.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368.

Lindblom, B. (1986). On the origin and purpose of discreteness and invariance in sound patterns. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 493–510). Hillsdale, NJ: Erlbaum.

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica, 57*, 297–314.

Locke, J. (1986). Speech perception and the emergent lexicon. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 240–250). Cambridge University Press.

Locke, J., & Pearson, D. (1990). Linguistic significance of babbling—evidence from a tracheostomized infant. *Journal of Child Language, 17*, 1–16.

Massaro, D. W. (1998). *Perceiving talking faces*. Cambridge, MA: MIT Press.

McCune, L., & Vihman, M. M. (2002). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language & Hearing Research, 44*, 670–684.

Miranda, E. R. (2002). *Computer sound design: Synthesis techniques and programming* (2nd ed.). Oxford, UK: Focal Press.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child-Development, 59*, 441–449.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188–194.

Rizzolatti, G., Fadiga, L., Gallesea, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor action. *Cognitive Brain Research, 3*, 131–141.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*, 131–145.

Vihman, M. M. (1991). Ontogeny of phonetic gestures. In I. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 69–84). Hillsdale, NJ: Lawrence Erlbaum.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61–82.

Vihman, M. M. (2002). The role of mirror neurons in the ontogeny of speech. In M. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 305–314). Amsterdam: John Benjamins.

Vihman, M. M. & Nakai, S. (2003). Experimental evidence for an effect of vocal experience on infant speech perception. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 1017–1020). Barcelona.

Wallace, V., Menn, L., & Yoshinaga-Itano, C. (1998). Is babble the gateway to speech for all children? A longitudinal study of children who are deaf or hard of hearing. *Volta Review, 100*, 121–148.

Westermann, G., & Miranda, E. R. (2002). Modelling the development of mirror neurons for auditory-motor integration. *Journal of New Music Research, 31:4*, 367–375.