Modelling the Development of Mirror Neurons for Auditory-Motor Integration

Gert Westerman[1] and Eduardo Reck Miranda[2]

[1]Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, University of London, London, United Kingdom, and [2]Centre for Theoretical and Computational Neuroscience, and Plymouth University School of Computing, Plymouth, United Kingdom

Abstract

In this paper we present a model of the integration of motor commands with their associated perceived sounds in vocal production. The model is neurobiologically plausible and is inspired by the development of vowels in an infant's babbling phase, wherein perceptual and action prototypes develop concurrently. When the model is exposed to external sounds, the perceptual prototypes develop culturally specific categories that coincide with these sounds. The model develops motor mirror neurons that are activated when their associated sounds are perceived and it learns to imitate sounds based on its own production categories.

1. Introduction

In order to introduce the problem addressed in this paper, let us imagine the following scenario: a pianist plays a melody on the piano and a violinist replays this melody on a violin. This trivial task poses a difficult Artificial Intelligence question: How can the violinist predict her movements on the violin in order to replay the melody that she heard on the piano? In other words, how does the brain associate sensory information with motor control? Suppose that the violinist is a robot who must learn how to accomplish this apparently trivial task. How can we program this robot's brain?

Miranda (2002b) has recently proposed an adaptive mechanism whereby interacting software agents are able to establish such associations symbolically. In this case, it is assumed that the agents have a brain capable of performing the subsymbolic neural tasks that are required for these associations. In this paper we propose a neural model that is capable of integrating motor commands and their associated sounds. In the experiments below we use the human voice, instead of a piano and violin.

Recent research has shown that in the categorization of stimuli, sensory domains are integrated at early stages of processing. This effect has most famously been investigated in speech perception, where visual stimuli (the face of the speaker) and auditory stimuli (the sounds produced by the speaker) are integrated in the listener's brain to form a unified percept. When the usual correlations between the visual and the auditory signals are experimentally altered by combining e.g., a visual /ga/ with an auditory /ba/, the speaker perceives a /da/ (McGurk & MacDonald, 1976). Westerman (2001) developed a model that was able to account for the integration of two sensory stimuli and the resulting perceptual changes in each domain. The model showed the development of prototypical perception in both domains as well as categorical perception and interference between domains based on inconsistent stimuli.

In this paper, we extend the model of sensory-sensory integration to one of sensory-motor integration. The questions that are addressed in this model are:

- What effect does sensory-motor integration have on the representation of both sensory stimuli and motor programs?
- How can these changes account for phenomena found in the development of infant's auditory perception and production?
- How can this integration serve as a basis for the imitation of perceived sounds?

Our model argues for a close link between perceptual and motor representations, a view that has recently gained much

Correspondence: E.R. Miranda, School of Computing, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK. E-mail: eduardo.miranda@plymouth.ac.uk

Motor Map   **Sensory Map**

Hebbian
connections
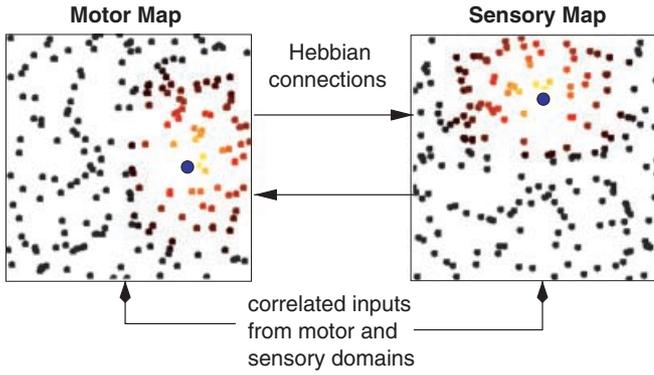

correlated inputs
from motor and
sensory domains

Fig. 1.   The basic architecture of the model.

popularity (Prinz, 1997; Hommel et al., in press). Further-more, based on the integration of action and perception, our model suggests an explanation of the origin of mirror neurons (Rizzolatti et al., 1996) for acoustic stimuli.

The rest of the paper is organized as follows: in the next section, we introduce the neural model and its respective sensory and motor apparatuses. Then, we demonstrate how the model learns the coupling between vocal control parameters and the resulting vocal sounds, leading to repre-sentational changes in both motor parameters and sound per-ception. Next, we show how the model can imitate heard sounds and how mirror neurons for auditory stimuli develop. Then we discuss the effect of exposing the model to exter-nal sounds; that is, heard sounds that were not produced by the model itself. Finally, we complete the paper with a general discussion on the performance of our model.

2. The model

2.1. Integrated neural maps

Our model integrates motor stimuli and their sensory effects. Both motor parameters and sensory stimuli form the inputs to two separate neural maps (Fig. 1). Each neural map con-sists of a number of units that are randomly positioned in the input space (the input spaces for both maps in Figure 1 are two-dimensional to allow illustration). Each unit acts as a receptive field with a Gaussian activation function: units that are close to an input signal (a black circle in the figure) are highly activated (lighter colour in the figure), and activation decreases with distance from the signal.

Mathematically, when an external input is presented on the map, the Gaussian activation of the units is computed as:

$$act_{i_{ext}} = e^{\frac{\|x - pos_i\|}{\sigma^2}} \qquad (1)$$

where pos_i is the position of unit i, x is the input signal and σ is the standard deviation (width) of the Gaussian function. Each unit is connected by unidirectional Hebbian weights to all units of the map for the other domain (Arbib, 1995). The Hebbian activation of a unit is the product of the

weight value vector and the activation vector of the units on the other map:

$$act_{i_{hebb}} = \sum_k act_k w_{ik} \qquad (2)$$

where w_{ik} is the weight from an unit k on the other map to the current unit i on this map.

The total activation of a unit is computed by adding the activations from the external stimulus and those from the Hebbian connections with the other map:

$$act_i = \gamma_e act_{i_{ext}} + \gamma_h act_{i_{hebb}} \qquad (3)$$

where γ_e and γ_h are weighting parameters to control how much each partial activation contributes to the total activa-tion of the unit.

The activation update after the presentation of a pattern is synchronous for all units, and the activation values are scaled to a maximum of 1.0.

One input to a map will typically activate several units, and the response r_i to an input x, that is, how the neural map "perceives" that input, is computed in a population code: the response is the vector sum of the positions of all units, weighted by their activation values:

$$r_x = \frac{\sum_i act_i pos_i}{\sum_i act_i} \qquad (4)$$

where pos_i is the position of unit i, and act_i is its activa-tion. Such population codes have been found to play a role for example in the encoding of motor commands in the monkey cortex (Georgopoulos et al., 1988) where the direc-tion of arm reaching is accurately predicted by adding the direction vectors of direction-sensitive motor neurons, weighted by their firing rates.

The Hebbian connections between the maps have initial values of 0. They are updated with the co-variance learning rule (Sejnowski, 1977):

$$\Delta w_{ik} = \alpha \left(act_i - \overline{act_i}\right)\left(act_k - \overline{act_k}\right) \qquad (5)$$

where $\overline{act_i}$ and $\overline{act_k}$ are the average activation values of units i and k over a certain time interval (30 iterations in our simulations). This rule strengthens the connections between units when their activation values are positively correlated, weakens them when the activations are negatively correlated, and does not change the weights when the activations are uncorrelated.

The correlation-based weight update has the consequence that units that respond to stimuli that consistently co-vary between the sensory and the motor map become more highly activated due to the strengthening Hebbian weights: co-varying signals between the maps result in the same respond-ing units on both maps having co-varying activation values, and thus developing strong Hebbian connections. This results in such units not only receiving external, but also strong Hebbian activation, and thus becoming more active than
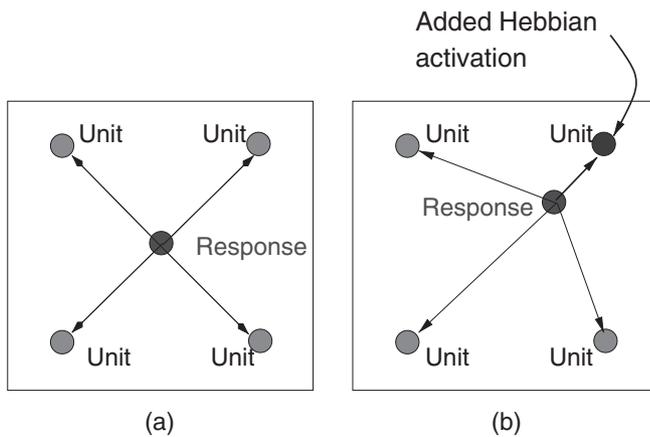
Fig. 2. The population-coded response to an input is influenced by external Hebbian activation. Without Hebbian activation, the response lies equidistant between activated units (a). When one unit is activated more due to Hebbian activation, the response is displaced towards that unit (b).

other units that do not reliably co-vary with units from the other map. Given that the population code response is weighted by unit activations, this means that such units "pull" the response towards them and induce a representational change (Fig. 2). Therefore, sensory/motor pairs that reliably co-vary will become more prototypical so that other, nearby inputs will be displaced towards them.

2.2. Sensory and motor apparatuses

Before presenting the results of the model, we briefly introduce here the implementation of the sensory and motor apparatuses that the model actually mediates.

2.2.1. The hearing model

The hearing apparatus senses audio in terms of formants. The formant analysis algorithm begins by re-sampling the sound to a sample rate of twice the value of the maximum formant frequency expected in the signal; an adult male speaker should have no formants higher than 5 kHz. Then the signal is filtered in order to increase its spectral slope. The filter is defined as follows:

$$\delta = e^{-2\pi F t} \qquad (6)$$

where F is the frequency above which the spectral slope will increase by 6 dB per octave and t is the sampling period of the sound. The filter works by changing each sample x_i of the sound, in reverse order: $x_i = x_i - \delta x_{i-1}$. Next, the signal is subjected to autoregression analysis (also known as predictive analysis). The target of autoregression is to create an all-zero filter that approximates the inverse of the acoustic resonator that originally shaped the spectrum of the sound in question. The analysis first reads snapshots of samples and estimates time-varying coefficients for an all-pole filter

that could recreate them. These estimated coefficients are inverted to fit an all-zero filter, which is applied to the signal in order to test the accuracy of the estimated parameters. In theory, the all-zero filter should cancel the effect of the resonator. The accuracy of the all-zero filter is measured by comparing its outcome with the original source signal; i.e., the result should be as similar as possible to the source signal prior to being shaped by the resonator. Each peak of the all-pole filter represents a formant in the spectral envelope and although a realistic all-pole filter response may contain a dozen peaks, only the resonating centre frequency of first two are considered for the purposes of our experiments. Autoregression analysis estimates the value s_t of a signal according to the following equation:

$$s_t = \sum_{i-1}^{p} (\phi_i s_{t-1}) - \omega_t \qquad (7)$$

where s_t is calculated by convolving the p number of predictive filter coefficients ϕ with the p known values of the signal, and ω_t is a white-noise driven signal that is filtered to yield a spectrum matching the input signal (Roads, 1986). In other words, the algorithm takes in several input samples and then uses the most recent sample as a reference to predict it from a sum of past samples weighted by the coefficient ϕ.

2.2.2. The vocal model

It is generally agreed that the vocal system can be effectively modelled as a resonating structure in the form of a complex pipe, on the condition that the model takes into account that the form of the pipe is variable (Miranda, 2002a). Also, the pipe should contain changeable internal obstructions, and the elastic properties of the walls of the pipe and its internal obstructions can change during sound production.

The synthesiser used in our experiments models the vocal system as a structure of chained short pipes, each representing a different section of the compound resonating system. Each of the pipes has four walls represented as Mass-Spring-Damper (MSD) units (Fig. 3).

Basically, the springs in Figure 3 represent muscles in the sense that their rest positions and tensions can be altered in order to adjust the walls and the internal obstructions of the pipe. The central idea here is that the walls and obstructions yield to air pressure changes and air is forced to flow inside the pipe as the result of mass inertia and elasticity. For example, by diminishing the volume of the lungs, air is displaced towards the mouth, acting as excitation for phonation. The model uses two types of equations to compute a sound: *myoelastic equations* that describe the physical behaviour of the walls of the pipes, and *aerodynamic equations* that describe the evolution of the movements and pressure of the air in the system (Boersma, 1995). The main myoelastic equation expresses the acceleration of one wall in the y-direction (height in Fig. 3) as follows:
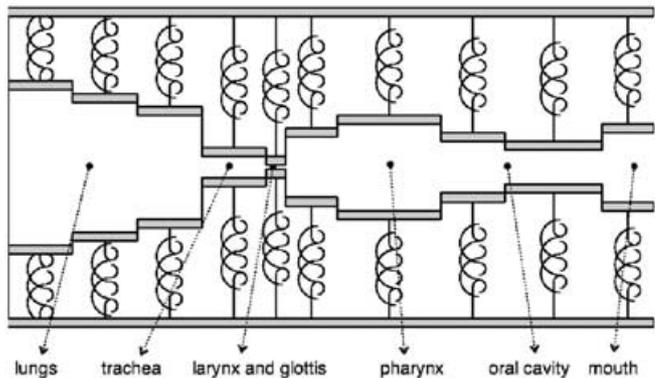
Fig. 3. Schematic representation of the vocal system. Each sub-pipe has four walls, but for the sake of clarity, the third dimension is not represented in this Figure. Also not represented here is the branching for the nose at the boundary between the pharynx and oral cavity.

$$m\frac{d^2 y}{dt^2} = F_T + F_C + F_D + F_A \qquad (8)$$

where m is the mass of the wall (in kg) and y is the displacement of the wall from the horizontal mid-line (in meters). The air pressure inside the pipe determines the development of the state of the pipe wall, which is represented by its displacement y and its velocity $\frac{dy}{dt}$. F_T is the tension force, that is the restoring spring force that brings the wall to its rest position; F_C is the collision force of the walls; F_D is the damping force that brings the velocity of the moving wall to zero and F_A is the air pressure force that pushes the walls apart when the pressure inside the tube is greater than the atmospheric pressure. The main aerodynamic equation is based upon the integral equation of continuity of mass flow between two arbitrary positions x_1 and x_2 (in meters) as follows:

$$\int_{x_1}^{x_2} A(x,t)\rho(x,t)dx \qquad (9)$$

where $A(x,t)$ expresses a position- and time-dependent cross section area (in m^2) of a pipe extending along the x-direction (length in Fig. 3) and $\rho(x,t)$ expresses the mass density of the air inside the pipe (in kg per m^3). A more detailed account of these equations is beyond the scope of this paper; please refer to (Boersma, 1998) for a thorough mathematical explanation. It suffices to say that the variables for these equations are specified by means of 29 parameters that metaphorically describe the actions of the vocal tract muscles and organs, such as cricothyroid, styloglossus, levator palatini, hyoglossus and orbicularis oris, to cite but a few (Fig. 4). For practical reasons, all parameters here are normalized to vary between 0.0 and 1.0.
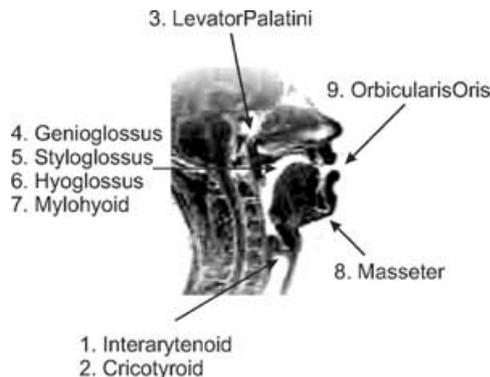

Fig. 4. The variables of the physical model voice synthesiser that are used in the examples and their approximate counterparts of the human vocal system.
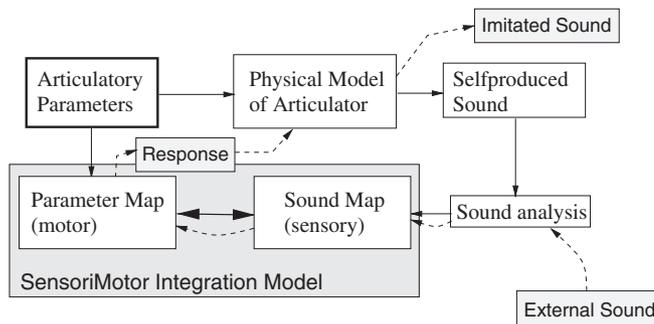

Fig. 5. The experimental setting of the model for babbling and learning the associations between the motor parameters and the resulting vowel sounds. The imitation pathway with the perception of an external sound is shown with dotted lines.

3. The experiments

Figure 5 shows the experimental setting of the model. In a babbling phase, the model randomly creates a set of articulatory (i.e., motor) parameters that form the input to the motor map; these are parameters used by the physical model voice synthesiser to synthesise the sounds. These sounds are then analysed with respect to their first two formant values, and these formant values form the input to the auditory map. Given the inputs to each map, the model is then trained as described above.

3.1. Representational change based on auditory-motor coupling

A first, simple experiment investigated the representational changes of motor and sensory stimuli as a consequence of the sensory-motor integration, and the imitation of heard sounds. For this purpose, two motor parameters, jaw opening and the position of the styloglossus muscle (a muscle that controls the position of the back of the tongue) were varied continuously at 18 steps each, and sounds were produced
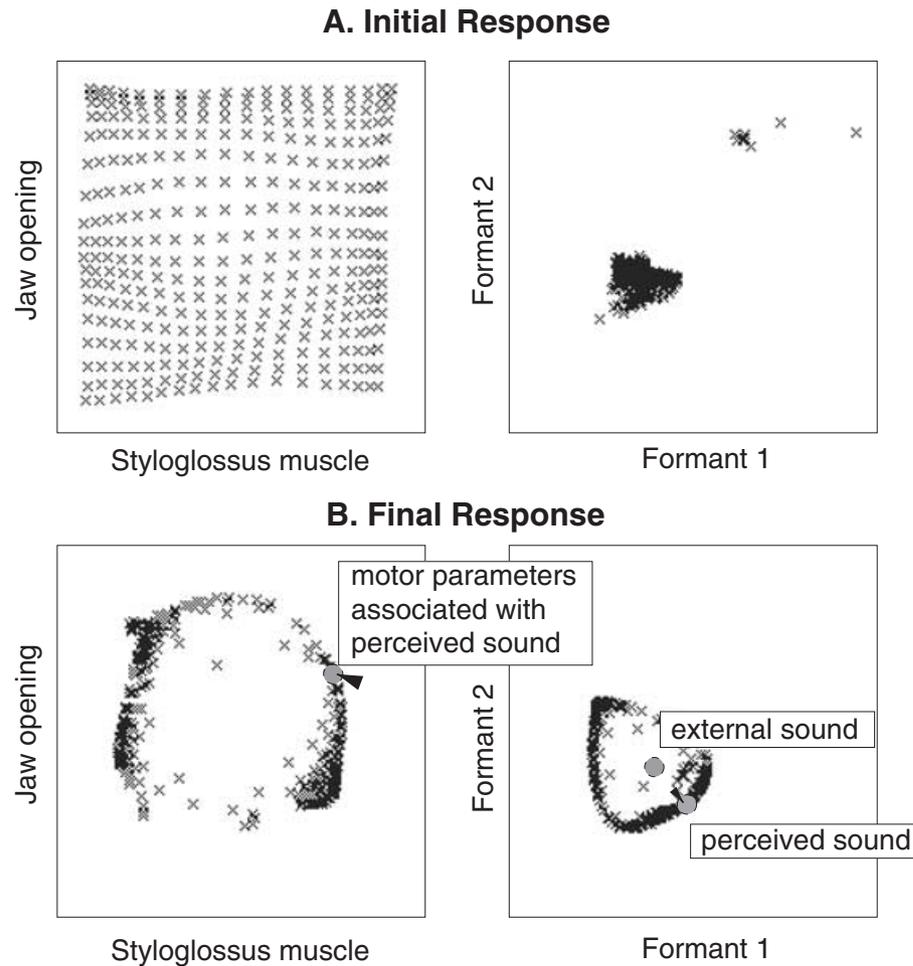
A. Initial Response

B. Final Response

Fig. 6. Initial (A) and final responses of the sensory-motor integration model (B). How an external sound is perceived and imitated in terms of a prototypical sound is also shown in B.

based on these parameters as described above. The model was trained on the resulting 324 motor-sound pairs. The parameters here were set to $\gamma_e = 1.0$, $\gamma_h = 0.4$ and $\alpha = 0.3$, for both maps.

Figure 6 shows the initial and final responses of the model. Initially all Hebbian connections were equal to zero, and the response of the maps is just the population code of the original data set: the motor map shows a continuous grid over the two parameters, and the sensory map shows the clusters of the resulting sounds. There is a non-linear relationship between motor parameters and resulting sounds: different motor parameter settings (e.g., with a high styloglossus muscle so that the tongue blocks the larynx) resulted in the same sound, or, in this case, in silence.

After the model has gone through a babbling phase by randomly choosing motor parameter sets and learning the associations between the motor and sensory parameters by adjusting the Hebbian connections between the maps, the response of the maps to the input signals looks very different (Fig. 6B). On both maps, prototypes in the form of dense clusters have formed concurrently, and preferred motor-

sound pairings have developed. The prototypes are formed around those motor-sound pairs that co-vary reliably. These are pairs where the mapping from parameter to sound is near-linear: a certain motor parameter set corresponds to a certain sound, and small variations in the parameters will only lead to small changes in the produced sound. This relationship will consistently co-activate the same units on the motor and auditory map and lead to strong weights between these units. By contrast, motor-sound pairs with highly non-linear relationships will not lead to strong weights: if small changes in the motor parameters cause big changes in the produced sounds, the motor units will be co-active with many different auditory neurons, and strong weights will not form. Therefore, motor prototypes form for those sounds that can be reliably produced; an aspect that also aids the perception of sounds by listeners.

This experiment shows the co-development of perceptual and motor representation and it illustrates how a perceptual event necessarily involves motor representations. Various studies support the notion that the perception of stimuli goes beyond the pure perceptual event in the sense that it seems

to involve motor representations. For example, it is well known that in the recognition of handwritten characters, subjects display knowledge of the dynamics necessary to produce these characters, and this knowledge can aid the recognition of the characters.

3.2. Imitation of sounds and the development of mirror neurons

Figure 6B shows how the model, after having learned the motor-perception coupling, can utilize this coupling to imitate sounds. An external sound that is presented to the model evokes a response on the auditory map. This response is propagated through the developed Hebbian connections to the motor map, causing a motor response that can be used to imitate the corresponding sound. However, the imitation of the heard sound is displaced toward one of the prototype clusters that have developed during training: the sound is perceived in terms of the prototypes that have been formed during the babbling phase (i.e., the population-coded response of the auditory map to the heard sound is displaced, indicated by an arrow on the auditory map in Fig. B). This way, imitation is not merely a reproduction of an external stimulus, but a re-interpretation of that stimulus based on the developed structure of the model.

The model suggests an account of mirror neurons that have, for example, been found in the pre-motor cortex of the monkey (Gallese et al., 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 the someone else. Mirror neurons also exist in humans, where the same neurons respond to grasping movements by the subject and the observation of such movements in others (Rizzolatti et al., 1996). In the model, mirror neurons develop naturally from the correlation-based coupling of actions and their perceptual consequences: a heard sound evokes a response on the motor map, and this response is the same as that needed to produce the sound itself.

3.3. External influences on the development of perception-action coupling

In a further experiment, we investigated the effect of an external environment on the development of the model. More specifically, we examined how the presence of sounds that are not produced by the model itself during the babbling phase shape its perceptual and motor properties. For this purpose, a more sophisticated set of motor parameters was generated by continuously varying six motor parameters. These parameters were:

a. the styloglossus muscle to control movement of the tongue backwards and upwards
b. the levator palatini muscle to control the raising of the velum

c. the hyoglossus muscle to pull the tongue body down and back for /a/-type vowels
d. the cricothyroid muscle to stretch the vocal chords
e. the interarytenoid muscle to adduct the vocal chords
f. lung pressure.

By this method, 1876 different motor parameter sets were generated. As in the previous experiment, for each set a sound was produced with the vocal synthesiser. The resulting 1876 sounds were again represented in terms of their first two formant values.

The external sound environment was created by segmenting 11 different vowels from French words spoken by a native French speaker, and encoding these vowels like the ones generated by the model, in terms of their first two formant values. Therefore, while the self-generated sounds existed in a motor parameter/formant pair, no motor-equivalent to the sounds existed for the external sounds.

The model was trained, as in the previous experiment, by randomly choosing a motor parameter set and presenting it and its associated sound to the model. This time, however, this babbling was randomly interspersed with external sounds, i.e., with the French vowels. For self-produced sounds, the parameters were set to $\gamma_e = 1.0$ and $\gamma_h = 0.4$ for both maps. When an external sound was presented and the motor map received no direct activation, γ_h for the motor map was set to 1.0. This approximated a mechanism by which direct activation of a map suppresses modulation by the other map $\alpha = 0.3$.

Figure 7 shows the responses on the auditory map before (A) and after (B) training without any externally perceived sounds. As in the previous experiment, before training, the response of the model is a faithful representation of the actual sounds. After training, the model has again formed densely clustered prototypes. The squares represent the French vowels, and without being exposed to these vowels during training, the model develops prototypes that depend solely on its non-linear mapping from motor parameters to sounds and that express the reliability of production. It is interesting to note, however, that many of the prototypes are very close to the actual French vowels.

This surprising result might indicate that in terms of language evolution, vowels in existing languages have developed based on perceptual-motor couplings, because they can be reliably produced. There may be two constraints at work in determining the vowel spaces of languages: the reliability with which the vowels can be produced, and the ability of the listener to identify them. Considering that the perception of vowels and the perception of "sound colour" (Slawson, 1985), or musical timbre, are fundamentally motivated by identical spectral principles (Handel, 1993), we can safely infer that this phenomenon is not limited to the realm of language but must play a key role in the evolution of musical perception.

In Figure 7B, it is only the "production" constraint that plays a role. This figure corresponds to a learner that is not

A. Initial Response
Formant 2 / Hz

B. Final Response without External Vowels
Formant 2 / Hz

C. Final Response with External Vowels
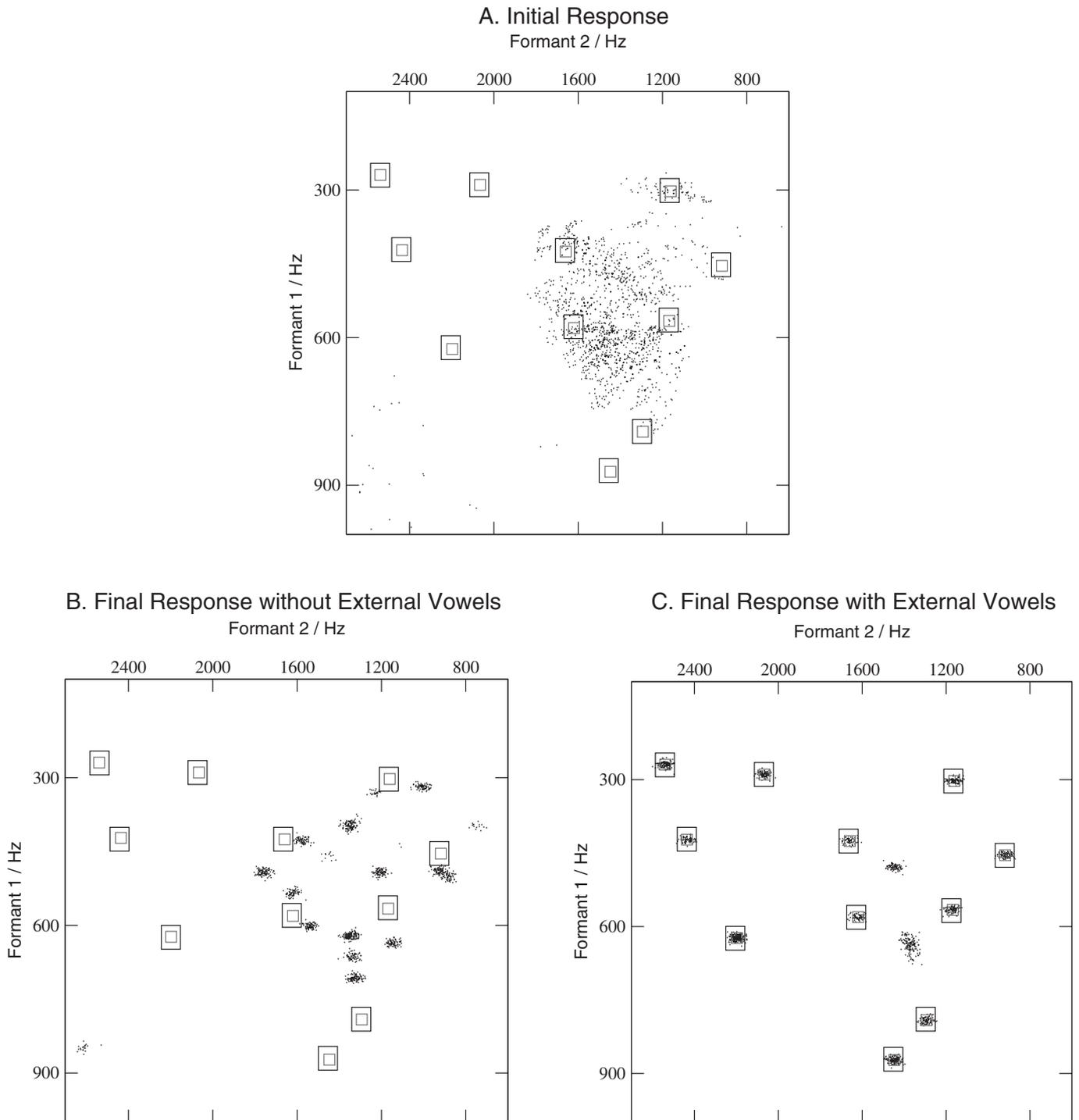### Formant 2 / Hz

Fig. 7. Auditory responses of the model when trained on 1876 motor-sound pairs. Each point represents a sound. The 11 French vowels are indicated by squares. Before training (A), and after training (B) without external vowels, and after training while being exposed to the externally produced French vowels (C).

exposed to sounds from others but instead only hears their own sounds. Therefore, perceptual prototypes develop solely based on the sensory-motor coupling. When the training of the model consisted of own-produced sounds interleaved with the French vowels, a different picture emerged: Figure C shows the auditory responses to the self-produced vowels

after the model was trained on these, but was simultaneously exposed to the French vowels (here, 90% of the training data were external vowels and the remaining 10% consisted of the self-generated motor-sound pairs). As a consequence, the perceptual responses to the majority of self-produced sounds now coincide with the French vowels.

How can the action of hearing an external sound that cannot be directly linked with its associated motor parameters lead to a representational shift in the perception-action coupling? In the spontaneously generated articulations of the model, when a motor parameter set is associated with the resulting sound, the Hebbian connections between co-varying units begin to strengthen. Then, when an external sound is heard but there is no external input to the motor map, this sound initially evokes a response only on the auditory map. However, as a result of the Hebbian connections, internal activation is generated on the motor map, and the motor units that normally co-vary with the activated auditory units also become activated. When the weights are then updated based on the covariance rule (Equation (5)), the connections between the now active units are further strengthened. In this way, hearing an external sound that was similarly produced before in babbling acts as positive feedback to the perception-action coupling for that sound. Therefore, the perceptual change and the learning of the ability to imitate a heard sound do not rely on its immediate imitation. This model introduces a significant improvement to the learning mechanism that we have been using in our simulations to study the evolution of music (Miranda, 2002b; 2002c), in the sense that here we take into account the importance of babbling for the formation of proto-perceptual categories, and we show that learning to imitate a perceived event can develop without explicitly performing that imitation.

4. Discussion

The model proposed in this paper captures several phenomena in the perception and production of sounds in a simple model. Firstly, it develops prototypical representations for motor commands and auditory stimuli. Further, it shows how the close perception-action coupling can lead to the development of the ability to imitate sounds and how this imitation is a re-interpretation of a perceived stimulus in terms of the developed system. Also, the model gives an account of how an external auditory environment can shape the perception of sound and it suggests how mirror neurons responding to auditory stimuli can develop.

The model is based on simple and neurobiologically plausible principles: receptive fields responding to a subset of all possible stimuli, population-coded representations of auditory stimuli and motor commands, and a simple Hebbian-based weight update. The Hebbian weights have initial values equal to 0, suggesting a role for activity-dependent structural growth in shaping cortical architectures (Quartz & Sejnowski, 1997). Further, the model suggests similar mechanisms in the integration between sensory stimuli and the integration between perception and action. Clearly, the model cannot yet capture very detailed aspects of the phenomena in question. Instead, we would like it to be understood as the outline of a suggested mechanism of the

integration between action and perception in a closely linked representational framework. A related such framework has recently been proposed by Hommel et al. (in press). They developed a theoretical framework of event coding (TEC) which postulates a common representational medium for perceptual contents and action plans. In our model, there is no explicit common code that unifies action and perception, but rather motor and perceptual representations modulate each other directly, via the connections between the domain maps. This view might be more in line with the lack of evidence for a neurophysiological substrate serving a common coding of perception and action (Decety & Grèzes, 1999).

Returning to our initial example of the violinist repeating the melody that was played on the piano, we can modify the model as follows: instead of matching vocal tract parameters with the produced sounds, the model can match the motor command for moving the fingers of the violinist with the sounds that are produced on the violin. In analogy with the imitation of speech sounds described here, a sound played on the piano will then activate the corresponding motor commands in the violinist, allowing her to re-play the heard sound or melody.

Acknowledgments

References

Arbib, M. (Ed.) (1995). *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge (MA).

Boersma, P. (1998). *Functional Phonology*. Doctoral thesis, University of Amsterdam, The Netherlands.

Boersma, P. (1995). Interaction between glottal and vocal-tract aerodynamics in a comprehensive model of the speech apparatus. In: *Proceedings of the International Congress of Phonetic Sciences* Vol. 2, pp. 430–433.

Decety, J. & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences* 3(5), pp. 172–178.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain 119*, pp. 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 – Coding of the direction of movement by a neural population. *Journal of Neuroscience 8*, pp. 2928–2937.

Handel, S. (1993). *Listening – An Introduction to the Perception of Auditory Events*. MIT Press, Cambridge (MA).

Hommel, B., Masseler, J., Aschersleben, G., & Prinz, W. (in press). The theory of event coding – A framework for perception and action planning. *Behavioral and Brain Sciences*.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264*, pp. 746–748.

Miranda, E.R. (2002a). *Computer Sound Design: Synthesis Techniques and Programming*, 2nd Edition. Focal Press, Oxford (UK).

Miranda, E.R. (2002b). Emergent sound repertoires in virtual societies. *Computer Music Journal 26*(2), pp. 81–94.

Miranda, E.R. (2002c). Mimetic Development of Intonation. In: Anagnostopoulou et al. (Eds.), *Proceedings of the Second International Conference on Music and Artificial Intelligence*, LNAI 2445, pp. 107–118. Springer, Berlin.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology 9*, pp. 129–154.

Quartz, S.R. & Sejnowski, T.J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences* 20, pp. 537–596.

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

Roads, C. (1996). *The Computer Music Tutorial*. MIT Press, Cambridge (MA).

Sejnowski, T.J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology* 4, pp. 303–312.

Slawson, W. (1985). *Sound Color*. University of California Press, Davis (CA).

Westerman, G. (2001). A model of perceptual change by domain integration. In: *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pp. 1100–1105.