Coupled Neural Maps for the Origins of Vowel Systems

Pierre-yves Oudeyer

Sony Computer Science Lab, Paris
py@csl.sony.fr

Abstract. A unified connectionist model of the perceptual magnet effect (the perceptual warping of vowels) is proposed, and relies on the concept of population coding in neural maps. Unlike what has been often stated, we claim that the imprecision of the classical "sum of vectors" coding/decoding scheme is not a drawback and can account for psychological observations. Furthermore, we show that coupling these neural maps allows the formation of vowel systems, which are shared symbolic systems, from initially continuous and uniform perception and production. This has important consequences for existing theories of phonetics.

1 Introduction

Our perception of vowel sounds is biased by our knowledge of our own language sound system. More precisely, Kuhl (1992) showed that when people are asked to evaluate the similarity between couples of vowel sounds, they tend to perceive two vowels closer than they are in a physical space when they are both close to a vowel prototype of their language, and further than they are in the same physical space when they are far from their vowel systems prototypes. In short, there is a perceptual warping that collapses percepts around the vowel prototypes of a given language. As a side effect, the perceptual difference between vowels belonging to different vowel categories is enhanced. This is refered in the literature as the "magnet effect". This is a particular instantiation of the well known and widely spread across modalities phenomenon of categorical perception, which Harnad (1987) defines as: "Categorical perception (CP) occurs when equal-sized physical differences in the signals arriving at our sensory receptors are perceived as smaller within categories and larger between categories".

Several connectionnists models of categorical perception, and in particular for the "magnet effect", have been developed. Among them, there are 2 broad categories: on the one hand, models like the Brain-State-in-a-Box of Anderson et al. (1977) or the backpropagation based model of Harnad (Harnad and Damper, 1997) which consist in training a feedforward neural network to reproduce its input point as outputs, with standard supervised learning training, in particular back-propagation ; on the other hand, models like the one of Guenther and Gadja (1996) which are based on the unsupervised training and self-organization of neural maps whose activity is interpreted with population codes as used in the pionneer work of Georgopoulos and colleagues (1988). Both kind of models have advantages and drawbacks.

G. Dorffner, H. Bischof, and K. Hornik (Eds.): ICANN 2001, LNCS 2130, pp. 1171–1176, 2001.

The first group of models provides more than the simple ability to simulate "magnet effect" like phenomena: it provides also the ability to categorize and identify stimuli, which allows to model on the one hand the ability to categorize, and also other aspects of categorical perception such as those concerning decision boundaries (Harnad et and Damper, 1997). The drawbacks of these models are that they are not very biologically plausible and need supervised training mechanisms (the BSB model is certainly more interesting from a cognitive modeling point of view than back-propagation training, especially for its modeling of categories as attractors, but it is brittle as Harnad and Damper 1997 explain). The model of Guenther does not have these drawbacks since there are many neurological evidences in favor of it, but looses the ability to categorize and identify stimuli. Moreover, no real comparison has been done so far to our knowledge.

In this paper, we propose to reformulate the model of Guenther in such a way that it can be interpreted in the framework of auto-associators, which allows a unified model of the "perceptual magnet effect", and extend it so as to provide the ability to categorize in a way similar to the BSB model. Moreover, we present a new experimental setup in which these neural maps are not used any more only to make models of the environmental vowels, but also for the production of vowels. Additionnally, instead of having one neural map that learns an already existing vowel system, we have many agents, each endowed with the neural system, that interact by producing sounds and listening to each other. They behave exactly as if they were actually learning an existing sound system, except that there is no such system (intially, each agent produces vowel that are spread uniformly across the space). We show that a phenomenon of self-organization arises in which all agents crystallise in a state where the distribution of vowels they produce is not uniform anymore, but comprises a number of sharply defined peaks: they did create effectively a sound system from scratch. We will explain why this has important implications for the current theories of the origins of communication systems, and in particular human language and sound system.

The next section describes the new formulation of Guenther's model, as well as its extensions. The following section describes the crystallization of neural maps when they are coupled. Then the result is replaced in a larger scientific framework in the conclusion.

2 A Neural Map for the Perception of Sounds

This model is based on topological neural maps, as Guenther. This type of neural network has been widely used for many models of cortical maps, which are the neural devices that humans have to represent aspects of the outside world (acoustic, visual, touch etc...). There are 2 neuroscientific findings on which our model relies, and that were initially made popular with the experiments of Georgopoulos (1988): on the one hand, for each neuron/receptive field in the map there exist a stimulus vector to which it responds maximally (and the response decreases when stimuli get further from this vector) ; on the other hand, from the set of activities of all neurons at a given moment one can predict the perceived stimulus or the motor output, by computing what is termed the population vector (see Georgopoulos 1988): it is the sum of all prefered vectors

of the neurons ponderated by their activity (normalized like here since we are intereseted in both direction and amplitude of the stimulus vector). When there are many neurons and the preferred vectors are uniformly spread across the space, the population vector corresponds accurately to the stimulus that gave rise to the activities of neurons, while when the distribution is inhomogeneous, some imprecisions appear. This imprecision has been the subjects of rich research, and many people proposed more precise variants (see Abbot and Salinas, 1996) to the formula of Georgopoulos because they assumed the sensory system coded exactly stimuli (and hence the formula of Georgopoulos must be somewhat false). On the contrary here we will show that this imprecision allows the interpretation of "magnet effect" like psychological phenomena, i.e. sensory illusions, and so may be a fundamental characteristic of neural maps.

Hence here a neural map consists of a set of neurons n_i whose "preferred" stimulus vector is noted v_i. The activity of neuron n_i when presented stimulus v is computed with a gaussian function: $act(n_i) = e^{-dist(v_i,v)^2/\sigma^2}$ (1) with sigma being a parameter of the simulation (to which it is very robust). The population vector is then: $pop(v) = \frac{\sum_i act(n_i)*v_i}{\sum_i act(n_i)}$ (2) The normalizing term is necessary here since we are not only interested in the direction of vectors. There are arguments for this being biologically acceptable (see Reggia 199?). Stimuli are here 2 dimensional, corresponding to the first two formants of vowel sounds. The difference with Guenther's model is that we use distances and gaussians while he uses scalar products and normalized vectors with agonist-antagonistic coding. Our formulation allows to unify this neural map model with auto-associators: indeed, (2) corresponds exactly to the Nadaraya-Watson estimator/regression formula (1964): $f(v) = \frac{\sum_i K(v,v_i)*f(v_i)}{\sum_i K(v,v_i)}$ (3) that is used to approximate some function f given a set of data points $(v_i, f(v_i))$, and where K is a kernel function. (2) can be mapped to (3) by taking a gaussian as the kernel function and $f(v_i) = v_i$, i.e. $f = Id$, which means that neural maps behave exactly like an auto-associator.

Initially, a neural map is formed by initializing the preferred vectors of neurons (i.e. their weights in a biological implementations) to random vectors. This is actually done through a babbling phase in a refined version of the model shortly described in next section. It means that the v_i are uniformly spread across the space. This can be visualized by plotting all the v_i as in one of the squares of figure 1. To visualize how agents initially perceive the vowel sound world in a way comparable to Kuhl's experiments, we can plot all the $pop(v)$ corresponding to a set of stimulus whose vectors values are the intersections of a regular grid covering the whole space. Figure 2 is an example of such an initial perceptual initial state: we see that the grid is nearly not deformed, which means as predicted by theoretical results, that the population vector is a rather accurate coding of the input space.

Then the learning mechanism used to updated these weights when presented a vowel sound stimulus v consists in shifting slightly these vectors towards the stimulus vectors when the v_i is close to it (activation is very high), shifting it away when it is in mid-range, and no shifting when it is far. In brief, this is a

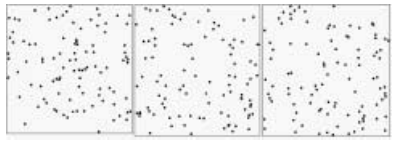

Fig. 1. Neural map at the beginning

sort of mexican hat based competitive learning mechanism. The actual formula is $\delta v_i = mexicanHatFunction(dist(v, v_i)) * (v - v_i)$. Furthermore, here a neural map represents an underlying vowel sound distribution, that one can compute by adding gaussians centered on all v_i's and then normalizing. Learning consists in adjusting neurons so that the coded distribution corresponds to the distribution of heard vowels.

When learning an existing sound system, agents are presented with vowel sounds coming from an existing vowel system (basically consisting of points grouped in as many clusters as there are vowels). As in experiments performed by Guenther, the kind of results we obtain are similar to what one can see on figures 4 (repartition of neurons) and 5 (image of a regular input grid), and fits perfectly well to experimental data found by Kuhl: the perceptual space has warped, i.e. input vowels close to the center of regions corresponding to a vowel of the language are perceive even closer and vice versa. This is due to the uneven repartition of neurons coupled with the imprecision of the population vector. A more precise way to visualize this is to plot the warping function with arrows giving the exact shift for a number of example points. Figure 7 is an example of warping function. What is interesting is that this looks like the plotting of attractor basins of dynamical systems. This leads to a possible refinement of the model that confers it the status of dynamical system and the ability to categorize/identify stimuli: as in the BSB model, the output of the neural map, $pop(v)$ can be re-directed to the input, thus making the map recurrent. The fact that this recurrent network has at the end effectively point attractors corresponding to prototypes for categories of vowels, which is suggested by figure 7, can easily be shown by casting formally the map onto a continuous hopfield network. This will be detailed in a longer paper.

3 Coupling Neural Maps

In the last section, neural maps were used only to perceive the sound world. One can imagine that they are also used in the process of vowel production (which is very sensible according to psychological evidence): more precisely, that the vowels produced by an agent at a given moment of its life follow the distribution coded by its perception map. This can be implemented in a biological plausile way by coupling the acoustic map with an articulatory/motor map, as will be detailed in a longer paper. Still, here we are only interested in the consequence of using perception maps to produce sounds.

The experiment presented consists in having a population of agents (typically 20 agents), each endowed with the neural system described previously. They interact by pairs of two (following the evolutionary cultural scheme devised in many models of the origins of language, see Steels 1997, Steels and

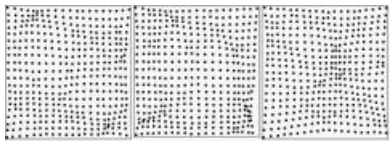

Fig. 2. Perceptual (non-)warping at the beginning (image of a regular grid of stimuli)

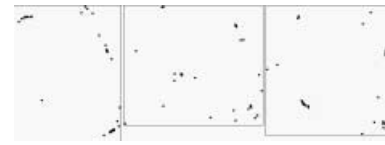

Fig. 3. Neural maps after 1000 interactions

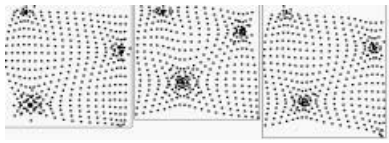

Fig. 4. Perceptual warping after 1000 interactions

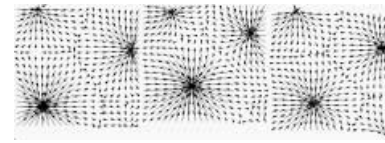

Fig. 5. Representation of the values of the warping at the edges of a regular gfrid

Oudeyer, 2000): at each round, one agent is choosen randomly and produces a sound according to the distribution coded by its map, and another agent is choosen randomly, hears the vowels and updates its map in the same way as above. Their maps are initially built by setting the v_i to random values, which means that the vowel they produce initially follow a quasi-uniform distribution. Figure 1 shows the initial state of neural maps for 3 agents in a simulation, figure 2 shows their initial perceptual warping (quasi non-existant). What is very interesting, is that this situation is not stable: rapidly, agents get in the a situation like on figures 3 and 4 which are the correspondances of figures 2 and 3 after 1000 interactions. Moreover, this final situation is experimentally stable, this is why we say that agents "crystalize". In fact, symmetry has broken and a positive feedback loop made that a distribution composed of a number of very well defined peaks appears, being the same for all agents (but different between 2 experiments). This final state is characteristic of the distribution of vowels in human languages. Moreover, because we added to the network the possibility to relax through recurrent connections, the peaks in vowel distributions are also attractors of the neural maps, and so good models of a categorizing behavior. In brief, this shows how a full fledged vowel system, which means not only a peaked distribution but also a shared system of discrete/symbolic units, can emerge through self-organization out of the coupling of unsupervised learning system.

4 Conclusion

The paper presented a connectionist model of 3 important phenomena: 1) the perceptual warping of vowels, due to the imprecision of population coding when receptive fields are not uniformly distributed ; 2) the emergence of a categorizing

behavior/discrete perception through the combined effect of imprecision and recurrent connections ; 3) the emergence of shared vowel systems with the coupling of unsupervised learning system, through symmetry breaking due to stochasticity and positive feedback loops. The last point is of particular importance to the field of linguistics. Indeed, many researchers (Archangeli and Langendoen 1997), defending nativist theories of language, think humans need to have in their genome many pre-specifications of phonemes, in particular vowel systems, and believe that biological evolution is responsible for their origins. For instance, (Kuhl 2000) proposed that we are innately given a number of vowel prototypes (all those that are "possible" in human languages), specified genetically, and that learning consists in pruning those that are not used in the environment. Our model shows that there is no need for linguistically specific devices. More importantly, nativist theories are faced with the problem of how these specifications got into the genes, to which they never provided some possible operational accounts. Our model shows that very generic neural devices can answer the questions, and need not biological evolution, but only cultural evolution (of course these neural devices were built through biological evolution, but their genericity indicates that language played certainly very little role for their selection).

References

Abbot L., Salinas E. (1994) Vector reconstruction from firing rates, Journal of computational Neuroscience, 1, 89-116.

Anderson J., Silverstein, Ritz, Jons (1977) Distinctive features, categorical perception and probability learning: some applications of a neural model, Psychologycal Review, 84, 413-451.

Watson G.S. (1964) Smooth regression analysis, Sankhya: The Indian Journal of Statistics. Series A, 26 359-372.

Reggia J.A., D'Autrechy C.L., Sutton G.G., Weinrich M. (1992) A competitive distribution theory of neocortical dynamics, Neural Computation, 4, 287-317.

R.I. Damper and S.R. Harnad (2000) Neural network modeling of categorical perception. Perception and Psychophysics, 62 p.843-867.

Georgopoulos, Kettner, Schwartz (1988), Primate motor cortex and free arm movement to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. Journal of Neurosciences, 8, pp. 2928-2937.

Guenther and Gjaja (1996) Magnet effect and neural maps, Journal of the Acoustical Society of America, vol. 100, pp. 1111-1121.

Kuhl, Williams, Lacerda, Stevens, Lindblom (1992), Linguistic experience alters phonetic perception in infants by 6 months of age. Science, 255, pp. 606-608.

Kuhl (2000) Language, mind and brain: experience alters perception, The New Cognitive Neurosciences, M. Gazzaniga (ed.), The MIT Press.

Steels, L. (1997a) The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1-35.

Steels L., Oudeyer P-y. (2000) The cultural evolution of phonological constraints in phonology, in Bedau, McCaskill, Packard and Rasmussen (eds.), Proceedings of the 7th International Conference on Artificial Life, pp. 382-391, MIT Press.

Archangeli and Langendoen (1997), Optimality theory, an overview, Blackwell Publishers.